DOCUMENT RESUME

ED 480 554 TM 035 177

AUTHOR	Jason, Martin H.
TITLE	Evaluating Programs To Increase Student Achievement: Leadership Strategies.
ISBN	ISBN-1-57517-815-X
PUB DATE	2003-00-00
NOTE	218p.; Foreword by Cozette Buckney.
AVAILABLE FROM	Pearson Skylight, 1900 East Lake Avenue, Glenview IL 60025 ($26.95). Tel: 800-348-4474 (Toll Free). Web site: http://www.skylightedu.com/ bkstore.
PUB TYPE	Books (010) -- Guides - Non-Classroom (055)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS
DESCRIPTORS	Elementary Secondary Education; *Evaluation Methods; Institutional Evaluation; *Program Evaluation

ABSTRACT

This book provides practical guidelines that personnel in an individual school can use in evaluating programs. It is fundamentally a book about winning. The current emphasis on test results as the single most important indicator of a school's effectiveness heightens the need for students, administrators, and staff to feel the pride that accompanies winning. Carrying out program evaluations can help meet this need. The focus is on analyzing program processes and student outcomes, whether academic or vocational. Investigating the processes used to deliver the program is essential, since they determine in large measure the impact of the curriculum and the attitudes of students and faculty. The first section, chapters 1 through 6 build a foundation of knowledge and attitudes applicable to conducting an evaluation study. The second section describes the phases for carrying out this evaluation through step-by-step procedures. The chapters are: (1) "Perspectives on Program Evaluation"; (2) "How Program Evaluation Contributes to a Learning Environment"; (3) "Using Program Evaluation To Improve the Curriculum—A Developmental Approach"; (4) "An Experimental Approach for Evaluating Programs"; (5) "Program Evaluation through Collaboration"; (6) "Measuring Program Outcomes"; (7) "The Evaluation Process: Phases 1,2, and 3"; (8) "The Evaluation Process: Phases 4, 5, and 6"; (9) "Writing the Evaluation Report"; (10) "Assessing the Evaluation Project"; and (11) "Revisiting the Principal's Leadership Role in Program Evaluation." Two appendixes discuss representative sources for authentic assessment and the ranges of effect size and corresponding percentile skills. (Contains 40 figures and 55 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

EVALUATING PROGRAMS

ED 480 554

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
C. Straka
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
2B

TO INCREASE STUDENT ACHIEVEMENT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM035177

MARTIN H. JASON

FOREWORD BY COLETTE BUCKNEY

BEST COPY AVAILABLE

EVALUATING PROGRAMS TO INCREASE STUDENT ACHIEVEMENT

LEADERSHIP STRATEGIES

MARTIN H. JASON, ED. D.

SkyLight Professional Development

Glenview, Illinois

Evaluating Programs to Increase Student Achievement: Leadership Strategies

Published by SkyLight Professional Development
1900 E. Lake Ave., Glenview, IL 60025
800-348-4474 or 847-657-7450
Fax 847-486-3813
info@skylightedu.com
http://www.skylightedu.com

© 2003 Pearson Education, Inc.
All rights reserved.
Printed in the United States of America.

LCCN 2002108661
ISBN 1-57517-815-X

V

Z Y X W V U T S R Q P O N M L K J I H G F E D C B A
10 09 08 07 06 05 04 03 02 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

There are

one-story intellects,

two-story intellects, and

three-story intellects with skylights.

All fact collectors, who have no aim beyond their facts, are

one-story minds.

Two-story minds

compare, reason, generalize,

using the labors of the fact collectors

as well as their own.

Three-story minds

idealize, imagine, predict—their best illumination comes from above,

through the skylight.

—Oliver Wendell Holmes

BEST COPY AVAILABLE

CONTENTS

Foreword xi
Acknowledgments xiii
Introduction xv

CHAPTER 1
PERSPECTIVES ON
PROGRAM EVALUATION

Overview 1
How Program Evaluation Contributes to School Improvement 1
Benefits of Site-Based Program Evaluation 4
Goals of Program Evaluation 5
Two Evaluation Methods 6
Assessing Program Processes 7
Linking Evaluation with Program Improvement 9
A Concluding Note 11
Chapter Highlights 13

CHAPTER 2
HOW PROGRAM EVALUATION CONTRIBUTES TO A LEARNING ORGANIZATION

Overview 17
The Spirit of a Learning Organization 17
The School as a Learning Organization 18
Reflective Practice 20
Linking the Parts of the School Organization for Improved Performance 23
Applying a Systems Approach to Increase Student Achievement 25
A Concluding Note 28
Chapter Highlights 29

CHAPTER 3
USING PROGRAM EVALUATION TO IMPROVE THE CURRICULUM— A DEVELOPMENTAL APPROACH

Overview 31
Compatible Evaluation Methods 31
The Problem of Confounding 34
Internal Validity 35
Some Further Threats to Internal Validity 39
A Concluding Note 41
Chapter Highlights 42

CHAPTER 4
AN EXPERIMENTAL APPROACH FOR EVALUATING PROGRAMS

Overview 45
Internal Validity—A Review 45
Building Control into an Experiment 46
Random Assignment 48
Experimental Designs 51
Deciding Whether to Use One or Two Groups in an Evaluation Study 55
External Validity 56
A Concluding Note 58
Chapter Highlights 59

CHAPTER 5
PROGRAM EVALUATION THROUGH COLLABORATION

Overview 61
Why Have an Evaluation Team? 61
Team Composition 63
Factors for Principals to Consider in Establishing the Evaluation Team 64
Forming an Evaluation Team 65
The Team in Operation 68
A Concluding Note 72
Chapter Highlights 74

CHAPTER 6
MEASURING PROGRAM OUTCOMES

Overview 77
Quantitative and Qualitative Data 77
Comprehensive Evaluation 78
Validity and Reliability of Evaluation Data 79
Validity 80
Reliability 85
Perspectives on Authentic Assessment 86
Focusing on Meaningful Learning and Creativity to Raise Test Performance 87
Linking Authentic and Traditional Assessment 89
A Concluding Note 90
Chapter Highlights 92

CHAPTER 7
THE EVALUATION PROCESS: PHASES 1, 2, AND 3

Overview 95
Phase 1—Describing the Program 96
Phase 2—Providing Direction for the Evaluation 99
Phase 3—Obtaining Information to Answer Evaluation Questions 103
Chapter Highlights 130

CHAPTER 8
THE EVALUATION PROCESS: PHASES 4, 5, AND 6

Overview 133
Phase 4—Analyzing Data to Assess a Program's Impact 133
Phase 5—Evaluating the Program's Effectiveness and Offering Recommendations for its Future Development 147
Phase 6—Writing the Evaluation Report 149
Chapter Highlights 150

CHAPTER 9
WRITING THE EVALUATION REPORT

Overview 151
Format and Content 152
What the Evaluation Report Represents 156
Chapter Highlights 161

CHAPTER 10
ASSESSING THE EVALUATION PROJECT

Overview 163
Purposes of Metaevaluation 163
Formative Metaevaluation 164
Summative Metaevaluation 168
A Concluding Note 171
Chapter Highlights 172

CHAPTER 11
REVISITING THE PRINCIPAL'S LEADERSHIP ROLE IN PROGRAM EVALUATION

Overview 173
The Multidimensional Responsibilities of the Principal in Program Evaluation 173
Attitudes of Principals Toward Assessing Programs 174
Beyond the Technical 176
Principals' Self-Assessment 178
A Concluding Note 178
Chapter Highlights 183

Appendix A: Representative Sources for Authentic Assessment 185
Appendix B: Ranges of Effect Size and Corresponding Percentile Shifts 187
Bibliography 191
Index 197

FOREWORD

In this era of site-based management, local control, and reform, the role, and importance of, the principal often becomes lost. What is the role of the school leader when "shared" with a parent council and empowered teaching staff? Leadership can, under those circumstances, become so diffuse as to have no real meaning. Effective school leadership requires that the goal of the school—educating students—be achieved. Accomplishing that task within the context of "shared leadership" is a challenge for leaders in the 21st century. That is why I am pleased to recommend Marty Jason's book on evaluating school programs.

The author addresses the current needs of principals as they strive to provide the teachers, and themselves, with the programs, assessment tools, and information needed to determine student progress. The book, beginning with the role that administrator can take in promoting, supporting, and facilitating the assessment process through the goals, process, and actual steps needed to review and evaluate programs, is a comprehensive guide for school leaders.

As a former elementary and high school principal, I often was in the position of determining whether a program or activity should be funded for the coming school year. My parent council, teachers, and students ultimately relied on my personal judgment and anecdotal information to make these decisions. Clearly, this was not the best way to determine whether to keep a program or activity.

As the former Chief Education Officer of a large inner city school district, it was important for me as well as my school leaders to make informed decisions. I can envision schools and school systems using this book to help them in this regard.

Jason's step-by-step evaluation process will be of great assistance to teachers, principals, and parent councils as they collaboratively make important decisions. Whether leaders need the foundation of knowledge and attitudes applicable to conducting an evaluation or a step-by-step procedure to begin an evaluation project, this book should be considered their guide.

Dr. Cozette Buckney
Former Liaison to the Board President/CEO,
Chicago Public Schools

Present Visiting Distinquished Professor of Urban Education,
Roosevelt University

ACKNOWLEDGMENTS

I would like to thank the following individuals who have helped me in various ways in writing this book:

Sue Schumer, former Acquisitions Editor at SkyLight Professional Development, for suggesting that I submit a proposal for this book based on a paper I presented at a conference. I appreciate your confidence in seeing the potential of this paper becoming a SkyLight publication, the warm support, useful information, and all of the work expended in my behalf. My best wishes for your future career.

Andrea Coens and Peggy Kulling, who in their capacity as project editors for SkyLight Professional Development, have provided valuable feedback and suggestions that have enhanced the writing of this work. You are a credit to your craft, and I very much appreciate your thoroughness.

I would also like to thank others at SkyLight who, in their various roles, aided in the production of this book. Their efforts are also very much appreciated.

Dr. Cozette Buckney, for taking time from your busy schedule while you were Chief of Staff for the Chicago Public Schools to write the foreword for this book. Since this work is designed primarily for practitioners, it is indeed gratifying to have an endorsement from a person who has occupied important leadership roles in the field. This type of feedback has provided crucial information that reinforces the scope and content of what this book has sought to accomplish. The fact that such feedback was given by a person of your reputation and level of professionalism makes it that much more valuable.

Dr. George Lowery, Dean of the College of Education at Roosevelt University, for continued interest in and recognition of not only my writing of this book, but my other scholarly activities as well. Your leadership exemplifies the type of support advocated in this work that would be gratifying to any staff member who believes he or she is engaged in a meaningful project. The genuine pleasure that you have shown in seeing one of your faculty members make a contribution to the field is a perk any professor would more than appreciate. Moreover, your delight in my accomplishments is testimony to the personal dimension so essential to administrator/staff relationships.

Martin Cohn, J.D., Attorney at Law, for the interest you have shown in my writing this book, your encouragement to do so, wise council given over many years, and not least of all, your valued friendship. I am very fortunate, indeed, to have known you for so long, have had your support, and experienced your sharp wit and keen insights brought to light during the many stimulating discussions we have had.

I would also like to thank the following individuals at Roosevelt University: Michael Gabriel, Reference Librarian, for his assistance in the ERIC search and David Green, Head of Reference and Information Services, for his perspectives on citing sources.

INTRODUCTION

Although this book provides practical guidelines that personnel in an individual school can follow in evaluating programs, it is fundamentally a book about winning. This perspective is consistent with a need identified by Sergiovanni (1984) for students and staff to "experience success [and] think of themselves as winners" (p. 13).

The current emphasis on test results as the single most important indicator of a school's effectiveness heightens the need for students, administrators, and staff to feel the pride that accompanies winning. Carrying out program evaluations can help meet this need since their ultimate aim is to determine how school outcomes can be improved. Even for those schools considered exemplary, assessing programs has the potential for taking student performance to still higher levels.

PURPOSE AND FORMAT OF THE BOOK

This book[1] provides guidelines that a team of faculty members can apply to evaluate ongoing programs that are either a regular or ancillary part of the curriculum. Moreover, the techniques presented can also be used to evaluate programs recently completed.

The focus is on analyzing program processes and student outcomes. The latter can include academic or vocational outcomes, or can concern development in other areas, such as improving self-concept or discipline. Investigating the processes used to deliver the program is essential since they determine in large measure the curriculum's impact as well as the attitudes of students and faculty

toward its instructional activities. In this context, evaluation of processes identifies specific factors operating at the core of the program that may be inhibiting successful outcomes. Such disclosure aids in generating solutions that could lead to improved student achievement. To this end, the methodology for obtaining, analyzing, and using "process" information is covered.

ORGANIZATION OF THE BOOK

This book is designed for practitioners with little or no experience in evaluating programs. Thus this activity offers an opportunity for further professional development. The content is divided into two sections. The first, consisting of chapters 1 through 6, builds a foundation of knowledge and attitudes applicable to conducting an evaluation study. The second section describes the phases for carrying out this assessment through step-by-step procedures that facilitate starting and completing an evaluation project. Examples are provided in order to clarify the various tasks involved in this process. In closing, chapter 11 highlights the importance of the principal's leadership in program evaluation.

Consistent with the subtitle of this book—Leadership Strategies—guidelines are presented that encompass perspectives and actions leaders can take to ensure that program evaluation as a tool for organization development has an opportunity to accomplish its purpose of providing direction for improving school outcomes. Therefore, how the leadership roles of the administration and faculty can be integrated synergistically with program evaluation and reflective practice to enhance the effectiveness and efficiency of a learning organization is also discussed.

Note that a distinction has been made between effectiveness and efficiency as criteria of organizational functioning. Effectiveness refers to the extent to which goals have been accomplished, whereas

efficiency considers the extent to which resources, financial and human, have been expended in accomplishing goals.

THRIVING IN AN ENVIRONMENT OF HIGH-STAKES ASSESSMENT

Decisions based on test results that may substantially or even adversely affect students as well as administrators and teachers characterize high stakes assessment (Jenkins 1993). Despite the growing criticism of using standardized tests in determining if schools are successful (e.g., Jenkins 1993; Kohn 2001; McColskey and McMunn 2000; Shepard 2000), these tests, at least in the foreseeable future, are not going to be discontinued.[2] One reason for this is that their results provide a uniform and relatively easily understood means of communicating to different groups the performance of a single student, a class, a school, or a district. Furthermore, findings from standardized testing provide a clear target to aim for when administrators and faculty seek improvement.

It is not the intent at this point to argue the benefits and limitations of standardized tests. In the final analysis, they constitute the reality that teachers and leaders at the building or district levels must deal with in this era of accountability. This should not be construed as negative. Helping students do well on standardized tests can serve as a challenge, a word not being used lightly here. Human experience teaches us that when individuals and groups confront difficult situations, they call on inner, untapped resources that enable them to meet these situations and grow stronger as a result. This means that educators who see only the downside of high-stakes assessment should instead regard standardized testing as an opportunity for growth.

The willingness of school personnel to assess their own work is an indicator of professionalism that can be expressed through program evaluation projects. In this regard, this book offers an approach to results-oriented inquiry that should pass the crucible of practicality.

A CONCLUDING NOTE

The importance of the leader's role in program evaluation is reiterated at various points in this book. With this orientation, the work describes the methodology to conduct a creditable evaluation. However, the principal needs to communicate to the staff that program evaluation is an integral part of the organizational fabric and follow up with leadership behaviors to actualize this priority. Otherwise, program evaluation projects are destined to become another school reform technique added to the heap of "tried but didn't work" initiatives. The call to leadership is heightened given this possibility.

ENDNOTES

1. This work is the outgrowth of a paper presented at the first Phi Delta Kappa International Conference on Effective Schools in 1999. The presentation involved a modification of materials the author has used with his graduate courses concerning program evaluation.
2. The No Child Left Behind Act of 2001 mandating standardized testing of math and reading in grades 3 through 8 supports this point.

CHAPTER ONE

PERSPECTIVES ON PROGRAM EVALUATION

OVERVIEW

No matter the area—academics, sports, music, or any other endeavor—in order to improve performance, current efforts *must* be evaluated. In this regard, program evaluation is indispensable for school improvement. Given this position, this chapter presents a case for local schools to conduct their own program evaluations, with a view toward raising student achievement. Also covered are the goals of program evaluation, how formative and summative evaluation methods complement each other, and the role of evaluation in program development.

HOW PROGRAM EVALUATION CONTRIBUTES TO SCHOOL IMPROVEMENT

Evaluation provides a tool for determining the extent to which a program[1] or curriculum is effective and at the same time indicates direction for remediating processes of the curriculum that do not contribute to successful student performance. Thus program evaluation serves two organizational functions—it confirms and it diagnoses (see Figure 1.1).

BEST COPY AVAILABLE

THE FUNCTIONS OF PROGRAM EVALUATION

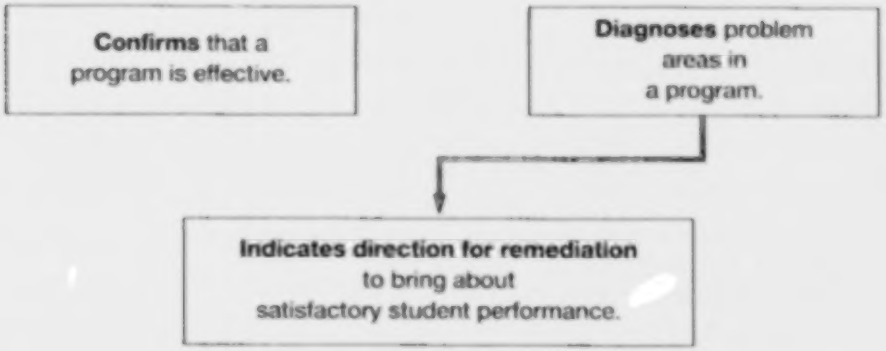

Confirms that a program is effective.
Diagnoses problem areas in a program.
Indicates direction for remediation to bring about satisfactory student performance.

Figure 1.1

Information disclosed by these functions enhances organizational efficiency by providing a focus for faculty and administrator efforts. This allows resources, in terms of time and money, to be directed at the areas of greatest need. As indicated in Figure 1.2, concentrating efforts where they are most needed helps ensure higher achievement levels. World-class athletes and superstars in the performing arts know this instinctively and work relentlessly to improve those aspects of their performance they know detract from their achieving an even higher level of excellence.

Concentrating efforts where they are most needed helps ensure higher achievement levels.

Educators can and should follow the example of superstars in other arenas. Delivering instructional activities is a performing art (Hasenstab and Wilson 1989) albeit one grounded in a science of pedagogical principles and concepts.

HOW PROGRAM EVALUATION CONTRIBUTES TO SCHOOL IMPROVEMENT

- Evaluation is a tool for determining effectiveness of a program or curriculum (confirms).
- Evaluation results indicate direction for remediating unsuccessful processes of the curriculum (diagnoses).
- Evaluation enhances organizational efficiency by providing a focus for faculty and administrator efforts.
- Enhanced organizational efficiency allows resources (time and money) to be directed at the areas of greatest need.

Figure 1.2

Principal and Faculty Leadership in Program Evaluation

For the principal and other school leaders who orchestrate the planning and implementing of programs, and for faculty who carry out learning activities, improving student performance is linked to improvement in processes that influence the effectiveness of teaching and learning. This approach is facilitated when the organizational structure of the school includes an evaluation team[2] that conducts program assessments on an ongoing basis.[3]

Against this background, program evaluation carried out by practitioners holds great promise for improving student achievement as reflected by test scores and other indicators. However, no evaluation project can reach its potential effectiveness without the leadership of the principal in promoting, supporting, and facilitating the assessment process. At the same time, no principal should take on the responsibility of managing an evaluation project. Doing so would take too much time from performing other administrative tasks, inhibit the professional development of staff members, and minimize or even eliminate the feeling of their ownership in the endeavor. Consequently, the satisfaction that accompanies the belief that one is part of some worthwhile activity would be lessened. When the

faculty has positive attitudes about their work, they produce self-sustaining energy because their efforts are likely to be gratifying, rather than being perceived as onerous.

Since the principal should not manage an evaluation project, there is a definite need for authentic faculty leadership to take the initiatives required for conducting an evaluation study as well as for carrying out the monitoring and follow-up of activities to ensure its success. To this end, the role of the leader of an evaluation team, along with the responsibilities of its members, are described in chapter 5.

BENEFITS OF SITE-BASED PROGRAM EVALUATION

Why should practitioners in a local school assess their own programs? This is a valid question given the pressures, time constraints, and stress faced by busy and often harried educators whose scrutiny of their work by the public, media, and legislators has translated into increasing demands for accountability. Criticism of schools for their culpability regarding substandard test performance has become a constant drone that is enervating for administrators and faculty alike.

In an environment where educators find themselves feeling defensive about their efforts and consequently working harder to meet and stave off further criticism, local school personnel are unlikely to relish taking on the additional responsibility of evaluating programs. This is understandable considering all they have to do to meet higher level expectations.

Given this situation, is conducting a systematic, ongoing program evaluation worth the time and effort? It definitely is, since this process functions as an organization development tool that can enhance professional growth and, correspondingly, student performance. Both goals would be accomplished because inherent in assessment projects is a guidance mechanism to facilitate identifying those aspects of the program's operation that may need modification. Moreover, by conducting their own evaluation of the extent to which a program impacts

student performance, educators demonstrate a willingness to be accountable to the outside community.[4] Figure 1.3 summarizes major advantages of each school assessing its own programs.

BENEFITS OF SITE-BASED PROGRAM EVALUATION

- Evaluation enhances professional growth.
- Evaluation results identify program weaknesses and lead to improvement.
- Improved programs enhance student performance.
- Educators involved in evaluation demonstrate a willingness to be accountable to the outside community.

Figure 1.3

GOALS OF PROGRAM EVALUATION

There are several aspects of program evaluation that are germane to the scope of this book. The following captures the intent of the content covered. For Brainard (1996),

> . . . effective program evaluation is a systematic process that focuses on program improvement and renewal and on discovering peaks of program excellence. Program evaluation should be viewed as an important ongoing activity, one that goes beyond research or simple fact-finding to inform decisions about the future shape of the program under study. (p. 5)

In this vein,

> . . . program evaluation contributes to quality services by providing feedback from program activities and outcomes to those who can make changes in programs or who decide which services are to be

> offered. Without feedback, human service programs (indeed, any activity) cannot be carried out effectively. (Posavac and Carey 1997, p. 14)

According to Smith and Glass (1987), based on evidence, value judgments are applied to determine a program's effectiveness. DeRoche (1987) provides a set of "principles of evaluation" relevant to the practitioner orientation of this book. Among them he maintains that the evaluation process should

> promote a positive attitude toward self-appraisal and self-improvement;
> provide opportunities for school personnel to diagnose difficulties;
> strengthen existing programs and establish pilot programs or projects to test new approaches;
> help teachers and learners determine the extent to which each has been successful in the teaching-learning process;
> encourage a team effort, a cooperative spirit, and a feeling by the community that all are accountable for the education of our young people. (p. 9)

Against the background of these authors' perspectives, conducting a site-based evaluation study fits within the scope of reflective practice. In this regard, an evaluation project offers a stimulus for administrators and staff to deliberate about their efforts and base subsequent actions on what they have analyzed.

TWO EVALUATION METHODS

Based on the work of Scriven, who in 1967 differentiated between two evaluation approaches—*formative* and *summative*—Worthen, Sanders, and Fitzpatrick (1997) describe these methods. They indicate that formative evaluation provides useful information for improving a program. In contrast, through summative evaluation, judgments are made about a "program's worth or merit in relation to important criteria" (p. 14).

Consistent with the guidelines presented in this book, Worthen et al. (1997) refer to a 1991 work by Scriven in which they indicate that "the results of [a] study may be used for both summative, and, later, formative evaluation purposes" (p. 16).

ASSESSING PROGRAM PROCESSES

While a summative evaluation will determine the extent to which a program has been successful, the key to improving school outcomes lies in analyzing how the program operates and making any needed adjustments. The reason for this is that the quality of the activities (processes) used to carry out a program can affect its outcomes (Weiss 1998). Thus summative evaluation can only go so far in suggesting areas of improvement since the information it provides is used to "summarize" evidence concerning the impact of the program. Whereas with formative evaluation, more specific ways to refine the processes (including materials) for delivering the curriculum are suggested. Put another way, summative evaulation provides a "barometer" by which indicators are measured to gauge how effectively a program has accomplished (or is accomplishing) its goals. But what also needs to be studied are the processes used in implementing instructional activities of the program.[5]

Formative Evaluation

As covered in chapter 6, there are also factors (e.g., teacher collaboration and students' preparedness for learning current content) that can affect instructional activities and can apply to *any* curriculum. Both curriculum processes and related factors provide the dynamics that determine the degree to which a program is functional. Studying these dynamics falls within the scope of formative evaluation since this type of assessment could disclose information leading to program improvement. Thus analyzing formative information provides the "how-to" dimension of program evaluation because, as indicated later in this book, opinions are elicited primarily from program staff on

actions to take to improve the processes and factors related to the delivery of a curriculum. In performance terms, formative evaluation provides the *focus* needed to concentrate on remediating those areas where less than optimum conditions inhibit higher levels of overall student proficiency. Therefore, formative evaluation contributes data to enable a program to be strengthened by guiding decisions on the direction(s) a program should take (Worthen et al. 1997).

This type of evaluation, according to Worthen et al. (1997), is useful when a program's development is in its early stages. Thus, before a program has been carried out too long, it can be improved by formative assessment. However, since outcomes depend on the processes used to achieve them, in keeping with the recommendation of these authors, a formative evaluation approach should be integrated with a summative study. Figure 1.4 highlights the compatibility of both assessment methods.

It should be noted that while formative evaluation can provide feedback (DeRoche 1987; Weiss 1998; Worthen et al. 1997) for improving the processes for delivering a program, such evaluation can also encompass analyzing interim indicators of its effectiveness, e.g., results of teacher-made tests. This kind of evidence could also indicate areas where modifications in the curriculum are needed and/or suggest the actions that should be taken to improve instructional activities.

Formative evaluation provides specific ways to refine the processes for delivering the curriculum.

While not diminishing the importance of summative evaluation, assessment of school programs should place heavy emphasis on formative methodology.[6] This perspective is in keeping with a defining characteristic of effective leaders in that they continually seek to improve problematic situations. Driving this outlook is their discontent with the status quo.[7] A corollary of this disposition is a leader's conviction communicated to the staff that "we can change what we need to." This belief has implications for professional development of the faculty and correspondingly enhances efforts in helping students reach their potential.

COMPLEMENTARY APPROACHES TO PROGRAM EVALUATION

	Formative	Summative
Purpose:	To improve a program's operation by examining its processes	To judge the worth, merit, or success of a program based on its effectiveness in meeting goals
When conducted:	In the early stages of a program's implementation and at the time a summative assessment is made	At the end of a program or agreed upon date after it has been in operation for a sufficient length of time to measure its impact
Data source:	Relies primarily on opinions from staff delivering the program on how it can be improved	Student outcomes, e.g., test and attitude scores; opinions of students, teachers, administrators, and parents

Figure 1.4

However, no matter how well this belief is conveyed, and even if it galvanizes a staff to work toward agreed-upon common goals, the enthusiasm for these will become short-lived without the mechanism for their attainment. In this regard, personnel in an individual school who are well versed in program evaluation methodology will be in a more knowledgeable and empowering position to effect change. Both formative and summative evaluation methods are related when efforts are directed toward program improvement as reflected in the following discussion.

LINKING EVALUATION WITH PROGRAM IMPROVEMENT

The process of improving a curriculum as described in this book is cyclical.[9] From this perspective, a current program is evaluated with respect to its outcomes (summative) and the processes that produced them (formative). Changes designed to improve instructional activities that are feasible to implement are incorporated into a revised

program. Next, the modified curriculum is tested experimentally and then reassessed summatively and formatively at an agreed-upon later date e.g., toward the end of the school year.

Applying these evaluation methods provides direction for improving the program. Thus program development is iterative. Each cycle leads to higher levels of curriculum quality, continually evolving in its content and delivery as program staff learn from their efforts. Figure 1.5 illustrates the program improvement cycle.

The price for excellence and beyond is continued effort.

CYCLICAL PROCESS OF IMPROVING A CURRICULUM

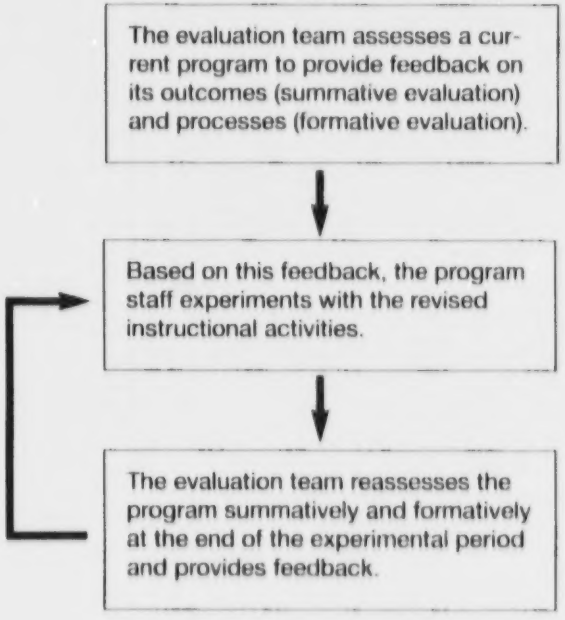

The evaluation team assesses a current program to provide feedback on its outcomes (summative evaluation) and processes (formative evaluation).
Based on this feedback, the program staff experiments with the revised instructional activities.
The evaluation team reassesses the program summatively and formatively at the end of the experimental period and provides feedback.

Figure 1.5

The entire process is grounded in a systems approach to organizational improvement. In this regard, the modified program is also referred to as an intervention since it represents a new factor introduced into the individual school as an open system.

The concept and dynamics of a systems approach to organizational performance is treated in chapter 2. At this point, however, it should be mentioned that an intervention also represents an innovative program since some of its features would be newly added. Finally, if it is decided that, based on an evaluation, a program ought to be discontinued and replaced with a more promising alternative, then this action can also be considered program development. The rationale for this is that the change represents the evolution of what may be a better curricular offering.

A CONCLUDING NOTE

The assessment approach taken in this book considers a program or a curriculum as a "work in progress" because those who deliver it should constantly seek to improve how its instructional activities are carried out, realizing that there could always be ways to make the process more effective and efficient. The notion of something being "in progress" does not imply that a lower standard is maintained until the program is finalized as a polished package, ready to be taught. No curriculum should be considered as ultimate.[10]

The price for excellence and beyond is continued effort. This is a good price to pay. This metaphor should be conveyed by leaders to their staff and, in turn, by faculty to their students because we do not know—and can never know—what can be finally achieved. To have such knowledge is a logical impossibility since the more we learn, the more our capacity to grow increases.

From an existential perspective, our existence as educators and the existence of those whose lives we touch in our roles can be immeasurably enhanced by unleashing the growth tendencies inherent in all human beings. "Unleashing" is an apt connotation because, under the

right circumstances, limits to one's learning become unfettered. Too idealistic? Too unrealistic? No! What *is* unrealistic are situations that prevent us from being who we really are and really can be.

Could a curriculum produce substantial change in students, not only from a knowledge perspective, but in a way that promotes their creativity? Without doubt it can, if the program is effective and efficient in accomplishing these goals. Thus, when meaningful learning activities are added to the equation, the dynamic for student and faculty growth is engaged.

In this vein, if the curriculum can be systematic in its plan and execution, and if program evaluation can be conducted systematically, then the confluence of both approaches should enable the continuous improvement of teaching and learning.

CHAPTER ONE HIGHLIGHTS

1. Program evaluation is indispensable for school improvement. It confirms the extent to which a program or curriculum is effective, and at the same time indicates direction for remediating processes of the curriculum that do not contribute to successful student performance.

2. No evaluation project can reach its potential without the leadership of the principal. However, a principal should not manage the project. There is a definite need for authentic faculty leadership to take the initiatives required for conducting an evaluation study as well as carrying out the monitoring and follow-up activities to ensure the project's success.

3. Conducting systematic, ongoing program evaluation is worth the time and effort since this process functions as an organization development tool that can enhance professional growth and, correspondingly, student performance.

4. By conducting their own evaluation of the extent to which a program impacts student achievement, evaluators demonstrate a willingness to be accountable to the outside community.

5. Summative evaluation determines the extent to which a program has accomplished or is accomplishing its goals. Formative evaluation provides information for improving the implementation of a program. The latter approach should be integrated into a summative study.

6. Evaluation of school programs should place heavy emphasis on formative methodology. This perspective is in keeping with a defining characteristic of effective leaders in that they continually seek to improve problematic situations.

7. Personnel in an individual school who are well versed in program evaluation methodology will be in a more knowledgeable and empowering position to effect change.

BEST COPY AVAILABLE

8. If the curriculum can be systematically planned and executed, and if program evaluation can also be conducted systematically, then the confluence of both approaches should enable the continuous improvement of instruction.

9. The process of improving a curriculum is cyclical. As such, a current program is analyzed with respect to its outcomes and the processes that produced them. Next, modifications are made that are designed to enhance instructional activities. The modified curriculum is then tested experimentally and reassessed summatively at a later date.

ENDNOTES

1. *Program* is a broad term that includes, but is not limited to, curricula (Gredler 1996).

2. DeRoche (1987) refers to this group as an evaluation committee rather than an evaluation team. Its function encompasses evaluating programs and other areas, e.g., faculty meetings and teacher effectiveness.

3. That program evaluation should be ongoing is consistent with what Brainard (1996) and DeRoche (1987) have advocated.

4. The use of evaluation to determine accountability is also cited in the literature (e.g., DeRoche 1987; Joint Committee on Standards for Educational Evaluation 1994; Posavac and Carey 1997; Weiss 1998).

5. According to Weiss (1998), how well a program teaches is one example of the processes that should be assessed by evaluators.

6. DeRoche (1987) favors formative as compared to summative evaluation. He points out that principals and teachers can make changes based on formative feedback during a program or project rather than waiting for its completion.

7. In discussing the school improvement process, DuFour and Eaker (1992) call attention to how innovation is blocked by accepting the status quo.

8. Other perspectives on the differences between formative and summative evaluation are also presented in figure form by Worthen et al. (1997).

9. DeRoche (1987) presents a ten-step curriculum improvement cycle that he relates to the evaluation process.

BEST COPY AVAILABLE

10. Weiss (1998) regards many programs as never finished since they continually change, not always due to improved strategies. However, she maintains that most often evaluation involves formative information to help staff improve their programs.

CHAPTER TWO

HOW PROGRAM EVALUATION CONTRIBUTES TO A LEARNING ORGANIZATION

OVERVIEW

This chapter describes the spirit that pervades the school as a learning organization. The connection of such an organization with program evaluation and reflective practice is also discussed. How the parts of the school organization can be linked by leadership attitudes and actions that promote viewing the school as a system is treated. From this perspective, various roles and functions of the school are viewed as interdependent, which thereby facilitates efforts directed at improving student achievement.

THE SPIRIT OF A LEARNING ORGANIZATION

Conducting a sound program evaluation extends beyond technique. It is part of an overriding mentality that those who aspire to higher levels of performance can readily embrace—the unremitting pursuit of excellence. This drive is captured in "*kaizen*, a Japanese word that connotes an ongoing spirit of concern with incremental but relentless improvement, however small" (Schmoker 1999, p. 51). Based on

BEST COPY AVAILABLE

McTighe's position advocating that schools adopt a spirit of *kaizen*, Schmoker argues that this spirit, with its emphasis on steady improvement, provides the power to combat fatalism in schools. He maintains that this can be achieved through, for example, feedback obtained from quarterly assessments, which indicates that "we are moving forward and becoming more capable" (Schmoker 1999, p. 52). Drawing on the work of Wiggins (1994), Schmoker also points out that small improvements should not be overlooked because to do so could affect teacher morale.

Adopting this spirit of *kaizen* does not necessarily limit us to small increments of growth. While it is essential for principals to recognize noticeable student and faculty accomplishments, principals should also use indicators of improvement as springboards to create a vision of even higher possibilities. Thus great things can be achieved when there is great vision. The words of Chicago architect Daniel Burnham reflect this poignantly: "Make no little plans, they have no magic to stir men's blood."

The anticipation of what students and faculty can accomplish can be realized when the curriculum is designed to optimize learning. Underlying this goal is the spirit of *kaizen*—a belief and feeling that pervades the individual school as an institution that is continually evolving because its members relate to the logical connection between their own growth and that of the school. This is the essence of the learning organization.

THE SCHOOL AS A LEARNING ORGANIZATION

Senge (1990) defines a *learning organization* as one "that is continually expanding its capacity to create its future" (p. 14). In terms of performance, he maintains that organizations that will truly excel will be those that have discovered "how to tap people's commitment and capacity to learn at *all* levels in an organization" (p. 4).

The notion that learning is evident at all levels in organizations that excel is philosophically compatible with the *transformational leadership* model. This may be noted from Burns' (1978) concept of transforming leadership[1]. He maintains that leaders and followers are both transformed because they are engaged in ways that mutually raise their motivation and morality and, consequently, their conduct and ethical aspirations.

Given the characteristics of a learning organization and transformational leadership, administrators who are open to their own potential for growth are likely to promote an environment that motivates and facilitates staff development. Moreover, collaboration between the principal and faculty to design a meaningful curriculum is, indeed, an ethical endeavor. Thus the choice of learning activities, by definition, affects the quality of school life for students as well as faculty. These activities should be perceived as worthwhile and desirable goals, since how individuals are required to spend time can elevate or lower the value of their experiences and consequently, their feelings toward the organization.

Program evaluation regularly carried out is a natural feature of a school as a learning organization. As such, its members can gain insights about which aspects of the curriculum produced the best learning conditions and which aspects need to be strengthened. In this regard, Posavac and Carey (1997) recommend that evaluators "encourage the development of a learning culture" (p. 262) in which staff and managers do not hide or condemn failures but instead learn from experiments.

The connection between this attitude and transformational leadership may be noted from a 1993 report of research conducted by Leithwood, Dart, Jantzi, and Steinbach (in Leithwood 1994). They found that one of the behaviors characteristic of principals displaying transformational leadership was their removal of penalties for making mistakes in professional and school improvement efforts. Consistent with the thrust of self-assessment of programs, these administrators also encouraged "staff to evaluate their practices and refine them as needed" (Leithwood 1994, p. 511).

Another behavior of the principal associated with the transformational leadership model developed by Leithwood (with various colleagues), involves having the staff "think reflectively and critically about their own practices" (Leithwood, Leonard, and Sharratt 1998, p. 249).

REFLECTIVE PRACTICE

Guiding the various phases of an evaluation project—from planning its design, to collecting and analyzing data, drawing conclusions, and offering recommendations for improvement—requires a level of systematic thought, both individual and collective (see Figure 2.1). In this vein, reflective practice is integral for assessing programs as suggested by the following definition based primarily on the work of Monti, York-Barr, Cronberg, Stevensen, Villejo, and Lunders (1998). Reflective practice is a cognitive process engaged in by one or more

PHASES OF AN EVALUATION PROJECT

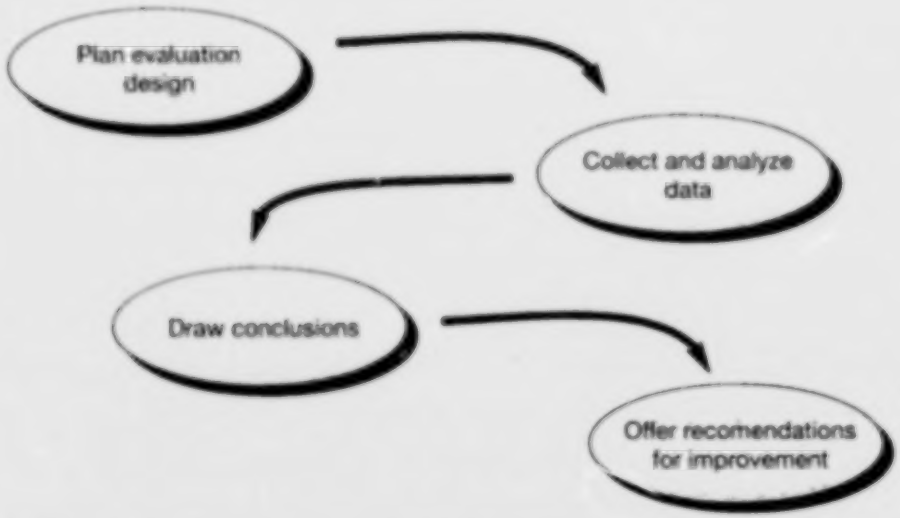


Figure 2.1

persons confronted by a problem situation that involves evaluation of values, goals, beliefs, policies, and practices, with a view toward implementing actions to improve the quality of school life for students and school personnel.

Given this definition, reflective practice is inherent in a learning organization. It serves as an overriding dynamic for personnel and institutional growth because progress depends on how we think about situations and, accordingly, how we act based on our conceptions.

Fundamentally, there are two dimensions to reflective thought—attitude and competency. In this vein, Senge (1990) introduces the concept of *reflective openness* and holds that it involves a "willingness to challenge our own thinking" (p. 277) as well as certain skills, such as team learning through dialogue and discussion.

Reflective practice is a cognitive process engaged in by one or more persons confronted by a problem situation that involves evaluation of values, goals, beliefs, policies, and practices, with a view of implementing actions to improve the quality of school life for students and school personnel.

Since engaging in reflective practice involves examining problem situations, the connection between this activity and the disposition of a learning organization may be noted from Leithwood et al. (1998). They point out that the stimulus for organizational learning "is assumed to be prompted by some felt need . . . or perception of a problem prompted from inside or outside the school that leads to a collective search for a solution" (p. 248). This search is facilitated through a program evaluation project since its activities are oriented systematically and comprehensively toward addressing a problem. In this context, problems are indispensable for personal and professional growth because their solutions often require creative approaches that can constitute curriculum modifications which in turn increase student motivation and learning.

As with anything else in life, it is how we choose to look at circumstances that defines our attitudes and actions concerning them.

Thus perceptions become our reality. From a phenomenological perspective, "all behavior is a function of the individual's perceptions" (Combs and Snygg 1959, p. 18). Put another way, to paraphrase Shakespeare: Nothing is bad, but thinking makes it so. The practical implications of this axiom are profound: Regard a problem as difficult, and it becomes so; regard a problem as an opportunity for improvement, and it becomes so. There is nothing unique about this statement—its message abounds in motivational literature.

That "problems should be considered opportunities" is not a trite maxim. It is imbedded in the mentality of a learning organization and is indispensable for systemic change. In this regard, promoting the attitude of determination and faith that improvements can and will be made, coupled with the confidence that we have the collective abilities to bring this about, provides a potent synergy that ripples throughout the school organization felt by staff and students alike.

Athletes and their coaches know well the potency of positive attitudes. It is the major factor that sustains them when fatigue sets in and it counteracts doubts and discouragement. Attitude also divides; it can differentiate world-class stars in any sport or performing art by the edge it gives the winner. In short, communicating a positive outlook toward school improvement efforts—from the initiative through implementation stages—*is the ultimate leadership strategy*. With a positive attitude, interventions are more likely to succeed. In contrast, a corollary of a negative attitude is lower expectations, and correspondingly, lower performance—grist for the frustration mill.

Communicating a positive outlook toward school improvement efforts—from the initiative through implementation stages—is the ultimate leadership strategy. With a positive attitude, interventions are more likely to succeed.

While a positive attitude motivates and sustains, it can only go so far. Thus attitudes without corresponding actions are short lived (see Figure 2.2). This situation can lead to cynicism on the part of the faculty who may come to believe that expressions of positive attitudes are just slogans, rhetoric, or exhortations that leaders are supposed to spout.

A structure for taking appropriate actions by the school as a learning organization is the evaluation team. The work of the team is greatly enhanced when the members of the organization at all levels adopt a systems mentality. This perspective is an overriding aspect of the learning organization as conceptualized by Senge (1990).

LINKING THE PARTS OF THE SCHOOL ORGANIZATION FOR IMPROVED PERFORMANCE

According to Senge (1990), "systems thinking is a discipline for seeing wholes. It is a framework for seeing interrelationships rather than things, for seeing patterns of change rather than 'snapshots'" (p. 68). For Senge, relying on snapshots means focusing on isolated parts of the system, with the result that the "deepest problems never seem to

THE POWER OF ATTITUDE ON ACTION

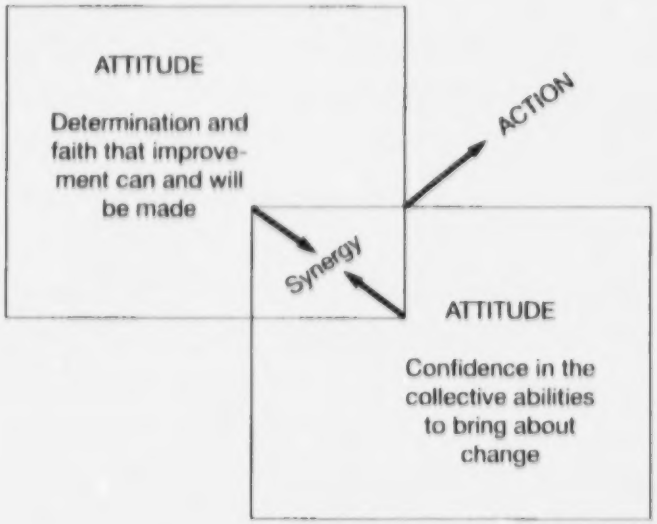


Figure 2.2

get solved" (p. 7). This malady can be prevented if we engage in systems thinking since Senge maintains that such thinking involves "leverage—seeing where actions and changes in structures can lead to significant, enduring improvements" (p. 114). Moreover, according to Senge, leverage also encompasses the notion that "small, well-focused actions can sometimes produce significant enduring improvements if they're in the right place" (p. 64).

When individuals engage in systems thinking, they can better understand the interconnectiveness of their work in the school organization.

When individuals engage in systems thinking, they can better understand the interconnectiveness of their work in the school organization. This insight facilitates higher organizational performance by bringing collective wisdom, produced through reflective practice, to bear on problems that prevent curricular processes from functioning optimally, thereby restricting student outcomes.

While theoretically systems thinking may be potentially effective and efficient, the physical arrangement of schools, with their compartmentalized classroom structure[2], is not designed to promote this perspective. In this regard, Schmoker (1999) cites Lortie's classic book, *Schoolteacher* (1975) when referring to the negative aspects of teachers' isolationism and individualism inherent in their work. This situation "thwarts . . . [teachers] from developing common solutions through dialogue. Isolation tacitly assumes that practitioners have nothing to learn from each other" (Schmoker 1999, p. 10). In stark contrast, Senge (1990) highlights the pivotal place that team learning has in a learning organization.

Schmoker (1999) also refers to Lortie's notion of presentism: the "myriad of daily events and duties" (Schmoker 1999, p. 10) that keeps teachers from reflecting collaboratively. However, as advocated by Schmoker and other writers cited by him, teamwork can counter this condition. This is supported by evidence he describes concerning the benefits of collaboration for students and teachers.

Individualism, isolationism, and presentism—exacerbated by the compartmentalized feature of schools—tend to inhibit the

communication process in collaborative improvement efforts. Moreover, by its very nature, teaching is not a "team sport." Yet, a common theme in school improvement literature regards collaboration as an essential factor in raising student achievement.

An implication of this theme for leadership is to provide a means to overcome obstacles to collaboration that would facilitate realizing the benefits of adopting a systems mentality. The provision for an evaluation team helps ensure that this outlook occurs. Thus in carrying out a program assessment, problems related to a lack of coordination regarding curricular processes among teachers can be identified. Information to remediate this deficiency can be obtained through a needs assessment. An instrument for this purpose is found in chapter 7.

By conducting program assessments, holistic thinking should become more pervasive if teachers experience the satisfaction of seeing higher student achievement as a consequence of linking instructional efforts across grade levels and/or disciplines. Furthermore, this experience is heightened when the staff believes that they are a part of a school that is special—one that is continually improving—and that all have a role to play in this endeavor.[3] To this end, the attitude and actions of the principal and evaluation team are indispensable for creating an environment in which the potential of a learning organization can become realized. Carrying out their respective roles serves as a driving force that propels and sustains the work of school improvement.

APPLYING A SYSTEMS APPROACH TO INCREASE STUDENT ACHIEVEMENT

Actions that reflect systems thinking can be taken to raise student performance in that such actions have schoolwide implications. With this orientation, attempts should be made to minimize the risks associated with high-stakes testing for students and school personnel. What is needed here is to change the mentality of such testing so that the stakes, while regarded as important, are not perceived to be so high that students' anxiety levels detract from their performance. In

short, mandated standardized testing should not take on a "do well or suffer the consequences" aura.[4] To prevent this counterproductive outlook, students need to be well prepared, beginning in the early grades, to perform competently on any exam. Underlying this goal is the belief on the part of students, teachers, and administrators that satisfactory (or even excellent) test scores are likely, rather than doubtful. The following recommendations may help develop this conviction:

- **Teachers across grade levels should hold continuous dialogues on ways to improve student achievement.**
  Teachers in the higher grades should communicate to those teaching lower grades the knowledge and skills they have found to be most needed for students to meet the requirements of the the higher grades successfully.

- **Teachers from the same grade level (or subject area) can exchange instructional techniques that they have found to be effective.**
  Teachers can also engage in collective problem solving, not only to enhance achievement, but to remedy other areas that can affect academic performance, such as behavior in and outside of the classroom. In this regard, McColskey and McMunn (2000) underscore the importance of collaboration for improving learning. They maintain that "in schools working toward instructional reform, administrators view teachers as professionals who need opportunities to work together to develop or fine-tune lessons, units, instructional materials, and assessment methods" (p. 119).

- **Students should begin to develop test-taking skills early.**
  These need to be taught throughout the grade levels.

- **Practice standardized tests should be administered.** However, the emphasis should not be on teaching the type of knowledge and skills likely to be on an upcoming "official" test, but to determine areas of strengths and weaknesses. When a pattern of deficiencies is disclosed, it can receive instructional focus and be covered in the curriculum by all of the teachers involved.

While the thrust of McColskey and McMunn's (2000) work argues for a long-term perspective of instructional reform, they also indicate what a school would "look like if educators focused exclusively on short-term fixes designed to boost test scores" (p. 116). Two of the short-term strategies they indicate are to administer practice tests and use computer programs "to enhance testing skills for students" (p. 117).[5]

The long-term strategy recommended by McColskey and McMunn (2000) of "providing students with meaningful and challenging work (not busy work)" (p. 118) is essential to developing individuals who are creative as well as effective problem solvers. However, the criterion for successful schools is still that which is most tangible to stakeholders—standardized test results. Thus educators need to survive the barrage of criticism leveled against lower performance in order to move in the direction advocated by McColskey and McMunn. In support of this point, McColskey, McCary, and Peel (as cited in McColskey and McMunn 2000) maintain that

> . . . while focusing on improving practice and quality will result in long-term improvement, short-term survival requires some attention be given to preparing students to perform on the type of state accountability test administered. However, addressing such a concern need not result in an exclusive preoccupation with raising test scores at any cost. Our approach has been to include not only test scores but also research-based instructional practices as the focus of dialogue with teachers and administrators to build a culture of inquiry. (p. 116)

Given this position, carrying out a program evaluation is congruous with the "culture of inquiry" inherent in a learning organization.

A CONCLUDING NOTE

Integrating formative and summative methods is consistent with reflective practice engaged in by members of a learning organization.[6] School personnel who function in this environment recognize that both types of assessment enhance their thinking about how student achievement can be raised. Thus formative and summative information stimulate dialogues concerning a) the reasons why certain instructional activities are more effective than others and b) ways in which future efforts of program delivery can be refined. These discussions bring an overriding sense of fulfillment when the learning organization moves inexorably forward, driven by the momentum of success that a systems approach to school improvement can engender.[7]

CHAPTER TWO HIGHLIGHTS

1. Program evaluation regularly carried out is a natural feature of a school as a learning organization because it provides a mechanism whereby its members can gain insights about which aspects of the curriculum produced the best learning conditions and which aspects need to be strengthened.

2. The various phases of an evaluation project—from planning its design, to collecting and analyzing data, drawing conclusions, and offering recommendations for improvement—requires a level of systematic thought.

3. Searching for a solution to a problem is facilitated through a program evaluation project. Its activities are oriented systematically and comprehensively toward addressing a problem.

4. Problems are indispensable for personal and professional growth because their solutions often require creative approaches, and these can constitute curriculum modifications that increase student motivation and learning.

5. A structure for taking appropriate actions by the school as a learning organization is the evaluation team. The work of the team is greatly enhanced when the members of the organization at all levels adopt a systems mentality. This perspective is an overriding aspect of the learning organization as conceptualized by Senge (1990).

6. Systems thinking (Senge 1990) is sound since individuals can better understand the interconnectiveness of their work in the school organization. This insight facilitates higher organizational performance by bringing collective wisdom, produced through reflective practice, to bear on problems that prevent curricular processes from functioning optimally, thereby restricting student outcomes.

7. In carrying out program assessment, problems related to a lack of coordination regarding curricular processes among teachers can be identified and remediated.

BEST COPY AVAILABLE

8. Students need to be well-prepared, beginning in the early grades, to perform competently on any exam. Underlying this goal is the belief on the part of students, teachers, and administrators that satisfactory (or even excellent) test scores are likely, rather than doubtful.

9. In an environment of a learning organization, school personnel recognize that, consistent with reflective practice, both formative and summative evaluation methods enhance their thinking about how student achievement can be raised.

ENDNOTES

1. Although Burns (1978) used the term *transforming leadership*, it is more commonly referred to as *transformational leadership*.

2. Based on Kanter's (1983) conception of companies that have segmentalist cultures, i.e., those characterized by compartmentalized structures, DuFour and Eaker (1992) indicate that this type of culture describes many school districts.

3. Sergiovanni (1999) points out the motivating effect that occurs when members of the school organization (including students) feel that they belong to something special.

4. The use of alternative assessment growing out of dire consequences that high-stakes testing could have for students, teachers, and administrators is discussed by Worthen et al. (1997).

5. These authors also raise questions about having every teacher overdo instruction of test-taking skills.

6. With respect to organizational learning, Weiss (1998) discusses one of the purposes of evaluation—that of providing feedback to enable the staff to reflect upon improving practice. From the work of Schön (1983), she indicates that a canny administrator may find that encouraging reflective practice is desirable.

7. Although not in the context of discussing school improvement, Csikszentmihalyi (1990) as cited in Schmoker (1999), maintains that the sense of moving forward as well as psychic satisfaction are provided by steady growth.

BEST COPY AVAILABLE

CHAPTER THREE

USING PROGRAM EVALUATION TO IMPROVE THE CURRICULUM— A DEVELOPMENTAL APPROACH

OVERVIEW

This chapter discusses how nonexperimental and experimental program evaluation provide a developmental and complementary approach to improve a curriculum. Based on the results obtained from nonexperimental research, a faculty can modify an existing program and assess changes. This calls for conducting program evaluation through experimentation. Accordingly, concepts and factors to consider in carrying out this type of investigation are covered. This information serves as background material for chapter 4, in which experimental methodology is treated in greater depth.

COMPATIBLE EVALUATION METHODS

Consistent with the growth concept of *kaizen*, evaluation carried out for the purpose of improving school achievement strengthens the link that connects the assessment results, the curriculum's ongoing development, and progressively higher student performance. In this vein, an ongoing program should be a dynamic entity, constantly evolving. Apropos to this notion is DuFour and Eaker's (1992) point that school

BEST COPY AVAILABLE

improvement efforts are endless. They would be, since achieving excellence in program outcomes involves a building process that, with each improvement, reveals discoveries of what students and teachers can accomplish.

Given this assumption, the quality of a program and its impact is enhanced when direction for its improvement is rooted in the best thinking of practitioners who have systematically and comprehensively tested their prescription for a more promising way to deliver instruction. To this end, program evaluation can be considered from two time perspectives—current and future. Each involves a different methodology that can be classified as either nonexperimental or experimental research.

Achieving excellence in program outcomes involves a continuous building process that, with each improvement, reveals discoveries of what students and teachers can accomplish.

Nonexperimental Program Evaluation

When assessing a program that is currently in operation (e.g., a regular curriculum), the time orientation is "present." That is, the program has already been implemented, and the investigators are now able to gather evidence regarding various dimensions of its effectiveness and related instructional processes.

This type of evaluation analyzes data with respect to judging the program *as it now stands*—prior to making any changes and is thereby classified under the umbrella of nonexperimental research. Such inquiry is important because it provides a rationale for introducing changes in the curriculum which are later tested experimentally. Thus both nonexperimental and experimental research are complementary from a curriculum development point of view.

Experimental Evaluation

In contrast to the nonexperimental method, a future time perspective is taken by conducting an experiment in which a modified or an innovative program[1] is evaluated to determine its impact. For this endeavor, a change based on consensus is introduced and, after waiting for a

period of time (e.g., one year), the extent of progress to date can be evaluated.[2]

Terms and Principles Associated with Experimental Evaluation

In carrying out an experiment, researchers *manipulate* the **independent variable**. This means that groups are treated differently; that is, each is provided with distinct learning activities. An independent variable consists of two or more treatments (or methods)[3], for example, different instructional programs, one of which could have a greater influence on the outcome than the other. Thus it is useful to make the connection between the **IN**dependent variable and **IN**fluence.

Students who receive the modified or innovative program comprise the experimental group, while those who are exposed to the regular program form the control group. The latter serves as the frame of reference for comparing the effectiveness of the experimental treatment or condition.

The **dependent variable** is also known as the **outcome variable** (or just **outcome**). This variable reflects differences in scores or other measures obtained from the groups. Thus the extent of these differences between the groups *depend* on the impact of a part of the independent variable, that is, a particular method or program. Put another way, the dependent variable also reveals the results of the treatments.

Nonexperimental inquiry is important because it provides a rationale for introducing changes in the curriculum.

Note that there can be an independent and dependent variable in nonexperimental as well as experimental research. However, the major characteristic that differentiates the two methods is the manipulation feature of the experimental approach.

The main advantage of experimental over nonexperimental research is that more control can be exercised over outside factors that can affect the outcome besides (or in addition to) the experimental treatment. (Features of both types of research as applied to program evaluation are shown in Figure 3.1.)

COMPARING EXPERIMENTAL AND NONEXPERIMENTAL METHODS OF PROGRAM EVALUATION

Feature	Experimental	Nonexperimental
Time orientation—Present		•
Time orientation—Future	•	
Manipulate the independent variable	•	
More control of extraneous variables	•	
Conduct formative and summative evaluation	•	•
Use of dependent (outcome) variables to measure results	•	•
Improve program based on results.	•	•

Figure 3.1

THE PROBLEM OF CONFOUNDING

When one or more outside factors have exerted influence, this undesirable occurrence is termed **confounding**. This situation is present when the combined influence of a treatment with one or more **extraneous variables** causes a noticeable difference in favor of the experimental group. These variables are considered extraneous because they are not part of the experiment. The problem here is that they could also affect the outcome. In fact, they can exert an even greater influence than a particular treatment. For example, students in the experimental group may have received more tutoring, or that group's teacher could have obtained additional professional knowledge during the course of the experiment. Such advantages could have boosted the achievement of students in the experimental group in addition to the treatment they received.

Under such circumstances, it would not be possible to determine how much of the differences (technically known as **variance**) could be attributed to the experimental treatment and how much could be

accounted for by extraneous factors. Thus confounding is present because there would be confusion when interpreting the results.[4]

Unplanned events such as those indicated may not be the only factors that could produce confounding. Other factors can also threaten the **internal validity** of experimental findings.

INTERNAL VALIDITY

This type of validity refers to a situation where it can be reasonably concluded that the experimental treatment was *solely* responsible for any differences and did not act in combination with a factor extraneous to the planned intervention, that is, the innovative program.[5] Unplanned, outside events (those not part of the experimental treatment) constitute unwanted influence and are labeled "history." It is one of a number of threats that lessen internal validity. More are examined later in this chapter.

History

This threat is so named because, during the course or **history** of the experiment, an event occurred that, together with a particular treatment, could have produced the outcome. For example, students in the experimental group may have gone on more field trips, heard invited speakers, or read additional interesting books recommended by their teacher. These activities could have increased student motivation.

At the time summative evidence is collected, it is not possible to eliminate the effect of history since the event has already taken place. If, however, an evaluation team becomes aware of an additional influence during the experimental period, this problem should be explained to those implementing the innovative program and a recommendation made that the unplanned learning activity not continue. If, for any reason, the experimental group teacher(s) would want to incorporate the ancillary approach as part of the program, then this change should be noted in the evaluation report. The reason for this is that if the experimental group is higher on all indicators, as compared

to the control group, the difference between them may not be accounted for solely by the teaching/learning activities of the intervention that were originally planned.

Rather than waiting until some event occurs before the faculty delivering the innovative program is asked to discontinue using the additional feature, it would be better for all to agree at the beginning of the experiment to implement the innovative program as it was designed.[6] On the other hand, during the course of the experiment, teachers in the experimental group may want to include other learning activities and materials that they feel can improve student achievement. If, in their professional judgment, these should be used in the future when the program is presented again, then this change should be encouraged.

When alterations are made, the faculty should provide the team with a written description of these. Modifications consistent with the goals of a program are in keeping with collecting formative data since the purpose of such data is to improve the operation of the program.

Not all unplanned events could motivate the program's faculty to make changes based on them. An extraneous factor over which the evaluators have no control could occur. For example, during the experimental period, a new department head may be appointed and bring to this role a vision of the program's potential, as well as providing enthusiastic support for it. This could have an energizing effect on those teaching in the program that may result in higher student outcomes.

There could be a situation in which an outside event could disadvantage the control group, thereby giving the experimental group an edge. For example, assume that there are two faculty members involved in an experiment. One teaches in the innovative program, the other in the regular curriculum; both have teacher aides. During the course of the experiment, a principal reassigns the aide from the regular curriculum—who happens to be experienced—to a new teacher who is not part of the experiment.

The replacement aide, though well intentioned, has not taken the initiative as did her predecessor, nor was she as skilled. As a result, the teacher has not been able to individualize instruction to the extent

that she did before because she has to do more follow-up of learning activities on her own. Thus the students have not received as much attention according to their deficiencies. In this situation, if the students in the experimental group out perform those in the regular program, the difference may not be a function of the intervention, but may be due to the control group being handicapped by not receiving as much individualized attention. Without this unexpected event of the change of aides, the achievement of the regular students may have been comparable to, or perhaps even surpassed, that of those in the innovative program.

Selection

Another threat to internal validity is referred to as **selection**. This describes the case where the innovative program has, for example, a sufficiently larger percentage of higher achieving and/or fewer problem behavior students at the beginning of the experiment. This imbalance gives a decided advantage to the innovative group if the outcomes are in its favor.

Attrition

The loss of participants in an experiment, known as **attrition**, threatens internal validity because these students are not available to provide summative assessment data. One reason for attrition is that participants could have moved from the school.[7] If both experimental and control groups are relatively small (e.g., an experiment involving only two classes), this threat could be a factor in accounting for any difference in, for example, test performance. Thus attrition could favor the experimental group if input cannot be included from lower achieving students in this group and/or if better achievers in the control group were not present to contribute summative information.

If both experimental and control groups are relatively small, attrition could be a factor in accounting for any difference in test performance.

Demoralization

Another situation that could be advantageous to students participating in the innovative program would be the **demoralization** of students in the regular program. It refers to a lower level of esprit for this group, affecting the motivation of its members, because they feel that they have not been selected to participate in the innovative program. The problem would be exacerbated if students in the control group were also in the same school as those in the experimental group.

Preventing or minimizing the potential for demoralization falls within the scope of the leadership role in program development. Thus the principal should be cautious about elevating the status and prestige of faculty members and students in the innovative program. Being overly laudatory about their activities could send an unintentional message to the control group that somehow its efforts are just ordinary—not exciting, state-of-the-art, or special. Moreover, making a group feel special in some way could also pose an internal validity threat since one would not be able to disentangle positive outcomes from the effects of boosts to the collective ego of the experimental group.

John Henry Effect

Instead of a group being demoralized because it was not selected to receive the innovative program, its teacher may regard the experimental class as rivals. Consequently, the control group teacher, in a competitive spirit, could strive harder to make his or her students perform even better than those receiving the intervention.[8] This feeling of rivalry and related behavior is known as the **John Henry effect** after the legendary railroad worker who competed against a machine. In this regard, if the compensatory efforts of the regular teacher are successful, they may mask the effectiveness of the innovative program.

Experimenter Effect

This effect can present another handicap for the intervention if those implementing the program do not carry it out as planned. An example would be modifying the time allowed for certain learning activities. In

contrast to the John Henry effect, the **experimenter effect** can work in *favor* of the intervention. In this situation, those implementing the innovation may exert extra effort in an attempt to have the program succeed. Another manifestation would be teacher enthusiasm—a potent and contagious factor in any learning situation. Its effect may be revealed in greater student motivation and subsequent higher achievement. Since the experimenter effect could combine with the activities of the program, it poses another threat to internal validity.

SOME FURTHER THREATS TO INTERNAL VALIDITY

An experiment can still be conducted even if it is not possible or practical to obtain a control group. This situation highlights the need for investigators to be particularly mindful of factors that could combine with the treatment if the experimental group shows noticeable improvement. Thus, in addition to threats already discussed, the presence of others can raise the issue of whether findings are valid. However, a control group provides the same opportunity for extraneous variables to affect student outcomes in both the intervention and regular programs. Put another way, factors not part of the experimental treatment are assumed to be balanced across both groups. This enables the intervention to be judged on its own merits. The following are additional threats to internal validity that should be considered, especially when using a one-group design.

Maturation

This refers to the normal growth of students over time as a factor contributing to the outcome in addition to that produced by the intervention.

Instrumentation

A change in the way students are measured on a posttest could facilitate their performance. For example, when compared to a pretest, if

the directions on a posttest are made clearer or some of its items revised, it may make it easier for students to demonstrate their knowledge and/or ability. This would then be reflected by higher test performance of the group.

Pretesting

Taking a test at the beginning of an experiment could alert students to the major topics that will be covered in the future. Believing that they will be tested again at a later date, the students may focus their studying on the various topics they have remembered from the pretest. This could boost performance on the posttest in addition to the learning activities of the innovative program. The threat of pretesting could therefore sensitize students to what is important to master. Pretesting could be a factor, especially for the more motivated students who could internalize the cues provided by the pretest. On the other hand, since students are routinely tested, pretesting may not be a concern. However, if the extent of improvement is being measured as an indicator of program effectiveness, then the evaluators have no choice but to include data from the pretest, whether or not one or two groups are involved in the experiment.

Another situation in which pretesting may be a factor is when a program extends over a relatively short period of time. To determine if pretesting has been a threat, students can be asked if they remembered things that were on the test that helped them learn new material.

A CONCLUDING NOTE

As stated earlier, confounding is present when the results of an intervention appear to be effective, but it cannot be stated with sufficient confidence that the success was due solely to the experimental program (see Figure 3.2). Thus threats to internal validity prevent conclusive statements that an intervention was effective. If these threats could be controlled, conclusions can then be drawn. Methods for achieving control in an experimental study are discussed in chapter 4.

VARIOUS THREATS TO INTERNAL VALIDITY THAT CAN BOOST AN INNOVATIVE PROGRAM'S EFFECTIVENESS

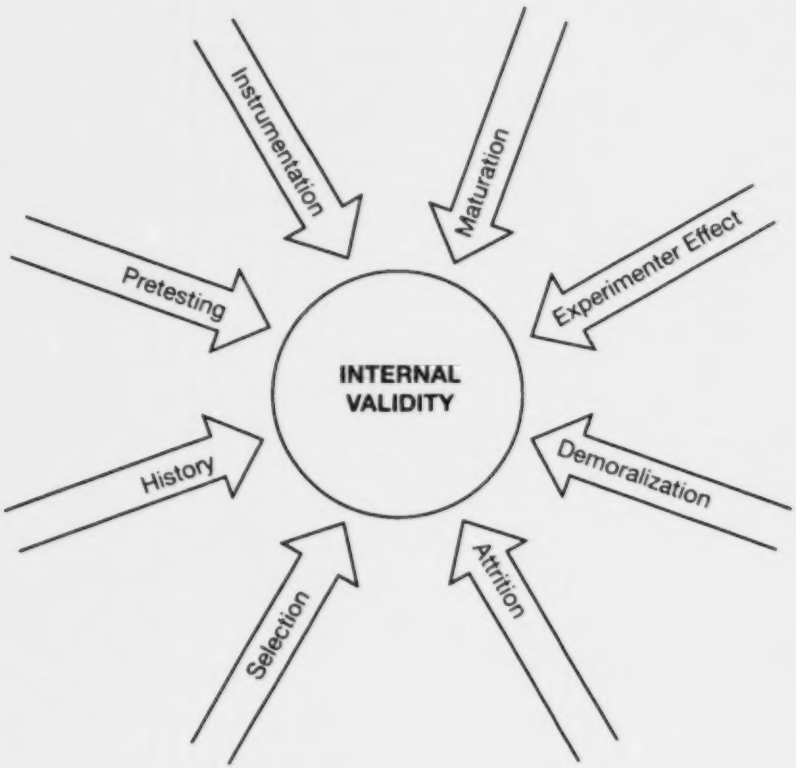


Figure 3.2

SkyLight Professional Development

CHAPTER THREE HIGHLIGHTS

1. Evaluation carried out for the purpose of improving school achievement strengthens the link that connects assessment results, the curriculum's ongoing development, and progressively higher student performance.

2. When the decision is made to assess a program that is currently in operation (e.g., a regular curriculum), then the time orientation is "present." This approach can be classified as nonexperimental program evaluation.

3. A future time perspective is taken by conducting an experiment in which a modified or an innovative program is evaluated to determine its impact.

4. In carrying out an experiment, researchers manipulate the **independent variable**. This means that groups are treated differently; that is, each is provided with distinct learning activities.

5. The innovative program is given to the students in the experimental group; those in the regular program constitute the control group.

6. The **dependent variable** is also known as the outcome variable. This variable reflects differences in scores or other measures obtained from the groups.

7. The main advantage of experimental over nonexperimental research is that more control can be exercised over outside factors (**extraneous variables**) that can affect the outcome besides (or in addition to) the experimental treatment.

8. If it is not possible to determine how much of the differences in an outcome variable could be attributed to the experimental treatment and how much could be accounted for by extraneous factors, **confounding** is present. This would lead to confusion when interpreting results.

BEST COPY AVAILABLE

9. **Internal validity** refers to a situation where it can be reasonably concluded that the experimental treatment was solely responsible for any differences and did not act in combination with any extraneous variables.

10. The following extraneous variables can boost an innovative program's effectiveness: history, selection, attrition, demoralization of students in the control group, experimenter effect, maturation, instrumentation, and pretesting.

11. Controlling threats to internal validity enables conclusions to be drawn about the effectiveness of the intervention with greater confidence.

ENDNOTES

1. The terms *modified* and *innovative program* as well as *intervention* are used synonymously in this book.

2. After the present program has been assessed, a decision may be made to experiment with an entirely new program, rather than modify the current offering. This action would still be consistent with taking a developmental perspective in seeking a potentially more effective curriculum.

3. While an independent variable can involve more than two treatments, in this book references are made to either a one- or two-group situation.

4. Suter (1998) also describes confounding as results that are confusing to interpret.

5. The notion of internal validity and how certain factors can threaten it was developed by Campbell and Stanley (1963). The material in this section discussing such factors is derived from their classic work.

6. The control group teacher(s) should also agree not to modify instruction. The reason for this is that altering learning activities for students in the control group changes the original comparison. In this regard, if the achievement of these students is higher as a result, the modifications could make the innovative program appear less effective.

7. Moving may not be the only factor. It is possible that students may not have taken the test(s) for a variety of reasons.

8. If there is more than one teacher in the regular program, the other instructor(s) could also act this way.

BEST COPY AVAILABLE

CHAPTER FOUR

AN EXPERIMENTAL APPROACH FOR EVALUATING PROGRAMS

OVERVIEW

This chapter extends the discussion of describing the methodology for conducting an experiment. First, the technical aspects of controlling extraneous variables are covered, followed by the use of different experimental designs. Issues in using a control group are also discussed. The chapter further points out the importance of being able to generalize the results of a program evaluation to apply to future students who would participate in the program.

INTERNAL VALIDITY—A REVIEW

Recall that when an experiment has internal validity this means that the experimental treatment itself (the innovative program) has produced a successful outcome unaided by one or more extraneous variables. Such variables involve factors that should not be part of any treatment given to the experimental and control groups and therefore must be controlled. If these factors are not controlled, this situation leads to confusion when interpreting results. Technically, confounding is present since one would not know if an outcome in favor of the

BEST COPY AVAILABLE

innovative program (e.g., higher test scores) could be attributed to the treatment it received, the extraneous variable(s), or a combination of both influences.

BUILDING CONTROL INTO AN EXPERIMENT

In order to control extraneous variables, nontreatment conditions (e.g., years of teaching experience) that can also have an effect on the outcomes are held constant. To achieve this, a control group is used. This provision enables the impact of nontreatment factors to be balanced across the innovative and regular programs. Therefore both groups have the same opportunity to be similarly affected by extraneous variables. The only variable that should not be held constant is, of course, the **independent** variable, i.e., the different programs being compared.

Using a control group enables the impact of nontreatment factors to be balanced across innovative and regular programs.

It should be pointed out that even with a control group, it is still possible that an unplanned event can take place that could benefit the experimental treatment. For example, as noted earlier, more students in the innovative program may have received tutoring than did those in the regular program. Except for this type of situation, having conditions that can affect both experimental and control groups in similar ways puts evaluators in a position to conclude more definitely that an innovative program was responsible for better results (if these in fact occurred) since nontreatment influences were controlled. Put another way, this conclusion can be drawn since factors that could affect outcomes besides the treatments were held constant for the experimental and control groups.

When it can be reasonably concluded that better student outcomes can be linked to the innovative, rather than the regular program, then the experiment can be considered as having internal validity. Thus the control group has served a valuable function by providing data as a frame of reference against which outcomes from the innovative program can be compared.

Planning an Experiment

To control for extraneous variables, think of the "level playing field" metaphor: Set up conditions so that the "game" can be played fairly. This cannot be done if the intervention has situational advantages over the regular program. To prevent this, the following factors should be as similar as possible for both the experimental and control groups at the beginning of the experiment:

- composition of the groups with respect to such factors as achievement level, behavioral problem students, socioeconomic level, and gender
- experience and qualifications of teachers
- time of day programs are offered
- physical environment of the classrooms
- material to be covered
- tests and other outcome measures to be used

Figure 4.1 synthesizes major points concerning the concept of internal validity. It is interesting to note how the prefix **CON** links these concepts.

Issues to consider when using a control group in an evaluation study are examined later. At this point, it would be useful to discuss random assignment—a control mechanism that can be applied to make the groups comparable.

ACHIEVING INTERNAL VALIDITY IN AN EXPERIMENT:
CONFIRMING IF THE EXPERIMENTAL TREATMENT ITSELF WAS EFFECTIVE

Results that cannot be disentangled from the influence of extraneous variables on an experimental treatment are confusing.

- **CON**fusing results are termed **CON**founding.
- Prevent **CON**founding by keeping nontreatment **CON**ditions **CON**stant for the experimental and control groups.
- When **CON**founding is not present, **CON**clusions about the effectiveness of the experimental treatment can be drawn more **CON**vincingly.

Figure 4.1

RANDOM ASSIGNMENT

When **random assignment** is used, the total number of subjects (students) that will participate in the experiment are divided to form the experimental and control groups. This process is based on chance, e.g., the flip of a coin, not the judgment of the person(s) responsible for planning the evaluation. The rationale for this is that an objective means is used to form the groups rather than a subjective opinion about who should be assigned to each program. The problem with relying on personal judgment is that, although not intentional, it may favor the experimental group.

Random assignment allows one to assume that both groups are equal (or nearly so) at the start of the experiment. This applies to any student variable that can affect the outcome besides the treatments e.g., prior achievement and/or motivational level. The basis for this assumption is that students who are high and low on certain characteristics will be balanced evenly between the groups. However, randomization *does not guarantee* that the students in both programs are comparable at the beginning. Even after the groups have been formed by this process, the regular program may still be disadvantaged because it has, for example, more students with a history of poor

classroom behavior. Conversely, the experimental group may hold the advantage if it contains a larger percentage of better students. Therefore, when more students with a certain characteristic are found in one of the groups, these particular students should be reassigned randomly so that the composition of the programs is balanced. Figure 4.2 summarizes what can and cannot be achieved by using random assignment.

The Use of Matching

Random assignment can be used in conjunction with **matching** to add further control since there would be greater assurance that students in both groups would be similar on a variable that could affect the outcome. Figure 4.3 lists the steps involved to integrate both processes.

RANDOM ASSIGNMENT: A METHOD OF EXPERIMENTAL CONTROL

RANDOM ASSIGNMENT

It does	*It does not*
form experimental and control groups objectively, by chance.	rely on subjective judgment to form groups.
assume that at the beginning of the experiment both groups are equal—or nearly so—on any student variable that can affect an outcome.	guarantee that the experimental and control groups will be comparable at the beginning of the experiment.

Figure 4.2

COMBINING RANDOM ASSIGNMENT WITH MATCHING

1. Use a measure that currently indicates students' achievement levels prior to the beginning of the experiment, e.g., the results of a standardized test.

2. Form two subgroups based on gender, and rank students from high to low in each one.

3. Taking one subgroup, pair the top two students, then the next two, and so on. Flip a coin; let heads and tails represent the experimental and control groups respectively. If heads appears, the first student in the pair is assigned to the experimental group; the other automatically goes to the control group. If tails, the first student goes to the control group and the other to the experimental group.

4. Repeat the process for each succeeding pair until there are no more left for that subgroup. Use the same method for the second subgroup.

5. If one student is left over from each subgroup, flip a coin to determine group assignment; do the same if there is only one student and no match.

Figure 4.3

Random assignment allows one to assume that both groups are equal (or nearly so) at the start of the experiment.

A Note on Random Assignment

While this process is designed to balance the groups, its use may not be feasible. Obviously, if the school year is in progress, classes cannot be re-formed for the sake of testing a program. The potential for negative reactions from parents, teachers, and students would not be worth it even if an innovative program appears promising. Thus random assignment is more practical when fall classes are being scheduled in the summer.

There are other limitations to randomization:

- In smaller schools there may be only one class at a particular grade level.
- The process cannot work if two schools are to separately represent the experimental and control groups.

Despite these limitations, random assignment can help control the possible influence of extraneous variables. However, its use raises an ethical issue since some students may not be the beneficiaries of a promising program (Worthen et al. 1997). From another perspective, Worthen et al. posed the question of whether it is ethical to expose people to a program that may be less effective than the current one.

Random assignment is more practical when fall classes are being scheduled in the summer.

It really does not matter by which means (random assignment or consensus) it is decided who will receive the intervention; the potential inequity of this determination is still a consideration. One approach to dealing with this dilemma in a school where a comparison is planned between an experimental and a regular program is to agree beforehand that if the intervention is successful, it (or its best features) will be adopted, if feasible, for students in the control group as well.

EXPERIMENTAL DESIGNS

These designs[1] can be thought of as a scheme for planning a study in which two groups or a single group will participate. The first design, shown in Figure 4.4, incorporates the use of random assignment (RA).

This design has three functions:

1. It measures the extent of improvement of each group over time. (Typically, the mean difference is used to determine change.)

RANDOMIZED PRETEST-POSTTEST CONTROL GROUP

	Group	Measurement 1	Treatment	Measurement 2
RA	Experimental	Pretest	Innovative program	Posttest
	Control	Pretest	Regular program	Posttest

Figure 4.4

2. The use of random assignment controls the threats of selection, maturation, pretesting, and instrumentation because each group has the same opportunity to be affected by these extraneous variables.
3. The pretest provides a check on whether groups are initially equivalent (or nearly so) on a variable that could affect the outcome in addition to the experimental treatment.

Recall that with random assignment one could assume that the groups were equal to begin with on any student factor, e.g., ability and attitudes, that could influence the results.

However, if it is not feasible to randomize participants since classes are already formed and if a control group is still to be used, then the design in Figure 4.5 would be appropriate.

NONRANDOMIZED PRETEST-POSTTEST CONTROL GROUP

Group	Measurement 1	Treatment	Measurement 2
Experimental	Pretest	Innovative program	Posttest
Control	Pretest	Regular program	Posttest

Figure 4.5

ONE-GROUP TIME SERIES

Pretest 1	Pretest 2	/ Intervention /	Posttest 1	Posttest 2

Figure 4.6

With this design one can only determine if groups are equal on the pretest and not assume that they are equal on other variables, as is the case when random assignment is used. Notwithstanding this caveat, the nonrandomized design still provides a practical plan for comparing the relative effectiveness of both treatments.

If only one group will participate in the experiment, due, for example, to ethical or practical reasons (Suter 1998), or if only one group is available, then the design shown in Figure 4.6 can be used.

The one-group time series design provides a way to determine the consistency of measures obtained before and after the intervention has taken place.[2] If, however, only one measure had been taken prior to introducing the intervention, student test performance may be lower than usual. This could be due to a number of reasons such as interruptions during the day, illness, or the test being administered by a substitute teacher. Lower achievement on a pretest can make the effect of an intervention appear more dramatic and thereby give a false impression of its impact. Thus having at least two measures can alert the evaluators to a marked discrepancy between them. If this occurs, the higher mean should be used. When the measures are close in value, the one nearest in time to the intervention provides the baseline data.

The advantage of a second posttest is that it measures the stability of the effects of the intervention. However, if only one posttest were given and scores declined for reasons such as those mentioned previously, this would mask the influence of the experimental treatment and lead to a conclusion biased against it.

As indicated in Figure 4.7, scheduling each pair of tests within a month seems reasonable. However, the decision regarding the

SAMPLE SCHEDULE FOR A ONE-GROUP TIME SERIES DESIGN

Pretest 1	Pretest 2	/ Intervention /	Posttest 1	Posttest 2
Early September	Late September		Early May	Late May

Figure 4.7

timeframe for testing should be made by the evaluation team and the program faculty. Moreover, alternative forms of the tests should be used, if possible, to minimize the opportunity for students to answer an item simply because they remembered it from the previous test.

To check on the durability of the treatment, follow-up testing could also be done in the fall, after the students have moved on to the next grade, since the program would not have been operating during the summer.

A Note on Using a One-Group Design

Since this design does not provide for including a comparison group, the only basis for judgment is the extent of growth from pre- to posttest. However, the change may also be due to the influence of one or more extraneous variables. It is, therefore, important to substantiate any evidence of noticeable improvement with the *judgment* of those who have implemented the innovative program. These faculty members, because of their everyday experiences with students, are in the best position to assess if marked progress has been made. They can express their opinions by answering **open-ended questions** that elicit more in-depth information from respondents concerning the program's impact. This is in contrast to using a format in which respondents are asked to indicate their opinions by checking a scale with several categories such as "very effective" through "not very effective." (How evaluation data can be collected by posing open-ended questions is discussed in chapter 7.) The following are

examples of the types of questions that should be used when a program involving only a single group is being assessed. (These are also appropriate for two-group designs.)

- How effective is the current program in raising reading achievement? Please explain.
- How would you compare the current program's effectiveness with the one previously used?

It may be the case that no students in the school are available to serve as a control group; or it may be decided that the program should be open to all those eligible because no one should be denied its potential benefits. Given these possibilities, a principal might be able to obtain the cooperation of another school similar in student population that would function as a control group. However, the use of a second school raises logistical issues with respect to such aspects as coordinating class content, tests, surveys with the staff, and scheduling experimental activities. Moreover, it may be difficult to find a school comparable on other factors besides student indicators e.g., teachers with experience similar to those in the innovative program, principal's leadership style, and school climate. Such factors are realities to consider when asking another school to cooperate. In this regard, the more that experimental and control schools are dissimilar, the more tentative should be the conclusions drawn regarding the intervention's effectiveness. Thus, while there is less control of extraneous variables in a one-group design, paradoxically, there is more control of the logistics of carrying out an evaluation when the assessment involves only one school.

DECIDING WHETHER TO USE ONE OR TWO GROUPS IN AN EVALUATION STUDY

While a two-group design controls conditions better and also provides a basis for a current comparison, this does not suggest that using only

one group would result in an unsound assessment. Even if a single group design raises one or more of the issues discussed earlier, the program should still be evaluated. However, as indicated previously, with one group the need to obtain corroborative evidence to provide a basis of comparison is underscored. Thus, if the students in the innovative program have characteristics similar to participants that experienced a different program the previous year, then their respective percentages of improvement can be compared.

It should be mentioned that analyzing a current group's performance with that of a previous one makes conclusions more tentative because conditions under which the programs were operating could have been markedly dissimilar, e.g., differences between teachers in years' experience and instructional style and differences in parental support and cooperation. This situation highlights the importance of having a teacher in the innovative program judge it relative to the one that he or she taught in the past.

As an aid to deciding whether to use one or two groups, features of the designs discussed are synthesized in Figure 4.8.

EXTERNAL VALIDITY

The extent to which a program has applicability for future students is another essential issue that needs to be taken into account in evaluating a program. This involves judging the extent to which other students similar to those participating in the intervention would benefit, provided, of course, that the treatment was found to be effective. Thus, if teachers believe that the students in the innovative program are typical of those who would participate in the future, then similar results can also be expected to occur for prospective students. Projecting the findings of the intervention to others refers to the experiment's **external validity** since conclusions can be extended, i.e., generalized, beyond the current investigation (Campbell and Stanley 1963).

FEATURES OF EXPERIMENTAL DESIGNS

Design	Used when	What it does
Randomized Pretest–Posttest Control Group	Students can be randomly assigned to experimental and control groups before school begins.	• Measures improvement • Random assignment controls for internal validity threats of selection, maturation, pretesting, and instrumentation • With random assignment it can be assumed that groups were equal on any student variable that can influence results • Pretest provides a check on whether groups are initially equivalent on a variable that could affect the outcome
Nonrandomized Pretest–Posttest Control Group	Intact classes representing the experimental and control groups are being compared.	• Measures improvement • Controls for the same threats as the randomized design • Determines if groups were equal on a pretest
One-Group Time Series	A comparison is not being made with a control group.	• Measures improvement • Additional pre- and posttests determine if measures obtained before and after the intervention has taken place were consistent

Figure 4.8

If, however, any situational factors, e.g., the threat of history, gave an advantage to the innovative program, external validity would be lessened. Such an occurrence would make the program less applicable for replication because the outside factor was not part of the program. A future frame of reference is congruent with the evolutionary aspect of program development. This lens conveys an optimistic perspective since it anticipates the promise that the program holds in terms of processes and outcomes for students as well as faculty.

A CONCLUDING NOTE

An experimental design serves as a framework for conducting the evaluation study. This framework provides an overall conception of procedures to be followed. Being aware of the "big picture" facilitates the work of the evaluation team because its members and other stakeholders whose input they depend on gain an understanding of how the evaluation will proceed. Moreover, describing the way the experimental design was applied should enhance the credibility of the evaluation report since this explanation reflects that a systematic approach was taken in assessing the program.

CHAPTER FOUR HIGHLIGHTS

1. Experimental designs are schemes for planning a study. The following designs measure the extent of improvement in program outcomes. The first two involve comparisons of two groups; the third applies to situations with a single group participating. Key features of each design follows:

   Randomized pretest-posttest control group

   Randomly assigning students to either the innovative or regular program assumes that both groups are equal (or nearly so) at the start of the experiment. This applies to any student variable that can affect the outcome besides the treatments.

   Pretesting provides a check on whether the groups are initially equivalent on a variable measuring student achievement.

   Nonrandomized pretest-posttest control group

   Use this design when it is not feasible to randomize, as in the case of already formed classes.

   Certain extraneous variables are controlled, but it cannot be assumed that both groups are initially equivalent on any student variable that could affect the outcome.

   One-group time series

   This design provides a way to determine the consistency of measures before and after the intervention has taken place.

   It does not control extraneous variables as in the previous designs.

2. Since control of extraneous variables is lacking in a one-group design, it is especially important to substantiate any evidence of noticeable improvement with the judgment of those who have implemented the program.

3. There may not be students available in the school who would be able to serve as a control group; or it may be decided that the

BEST COPY AVAILABLE

program should be open to all those eligible since no one should be denied its potential benefits.

4. Given these possibilities, a principal may be able to obtain the cooperation of another school similar in student population that would function as a control group. However, the use of this school raises logistical issues.

5. Projecting the findings of the intervention to other students refers to the experiment's external validity.

6. An experimental design serves as a framework for conducting an evaluation, thereby facilitating the work of the evaluation team. Describing the design should also enhance the credibility of the evaluation report.

ENDNOTES

1. The designs presented in this chapter are based on the work of Campbell and Stanley (1963).

2. Only two pre- and post-measures are shown here. However, if further checks are desired, other testings can be added.

BEST COPY AVAILABLE

CHAPTER FIVE

PROGRAM EVALUATION THROUGH COLLABORATION

OVERVIEW

This chapter discusses the advantages of conducting a program evaluation by a team composed of faculty members. The following are also covered:

- factors for principals to consider in establishing the evaluation team,
- the process for principals to follow in forming an evaluation team,
- guidelines for helping the team function effectively and efficiently,
- preliminary team activities, and
- suggestions for the team leader.

WHY HAVE AN EVALUATION TEAM?

There are several advantages to conducting program evaluation through a team approach:

- collective judgment,
- diverse base of knowledge and skills,
- varied perspectives, and
- shared responsibility.

These are discussed next.

BEST COPY AVAILABLE

Collective Judgment

In the final analysis, evaluation involves a judgment concerning the effectiveness of a program as well as future directions it should take. These opinions can be enhanced by capitalizing on the collective professional knowledge and insights team members can bring. Consistent with Smith (1996), a team represents a critical mass of individuals who can effect change. Their input can provide a synergy of ideas that can increase the effectiveness and efficiency of delivering a curriculum.

Diverse Base of Knowledge and Skills

Conducting a sound evaluation requires diverse competencies. These include knowledge and skills in the areas of:

- leadership,
- management,
- human relations,
- reflective practice,
- data analysis, and
- curriculum.[1]

Furthermore, as a professional development activity, an evaluation project provides a variety of experiences to augment these competencies.

Varied Perspectives

A team composed of faculty members from different grades and/or subjects provides the opportunity for varied perspectives to surface in judging the worth of a program.[2] This situation can lead to an assessment that is fair and perceived as such. This is not to suggest that only through a team approach can a fair program assessment be made. Certainly, an individual evaluator could do so. However, the point is that a team approach has the advantage of determining the effectiveness of a program when different points of view coalesce in reaching consensus. Thus the judgment of the team is supported by the weight of group members who, through dialogue and careful consideration of the evidence, have arrived at a common conclusion

regarding a program's strengths and also directions for its improvement.[3] Moreover, that this conclusion represents an assessment by peers who understand the conditions and school environment in which the program operates[4] adds to its fairness.

Shared Responsibility

A comprehensive evaluation project requires sufficient time and effort to be completed satisfactorily. While comprehensiveness can help withstand the potential criticism of not having adequate and representative evidence to make judgments, there are simply too many tasks for one or two people to handle in a timely fashion. Thus, a team approach makes the effort more manageable.

TEAM COMPOSITION

The team should represent teachers from other grades (or subjects), excluding those staff members implementing the program. This lends credibility to the evaluation report since it would not be written by those who have a vested interest in the program's success. The rationale for this broader representation is not only political: Having faculty with diverse teaching responsibilities participate in an evaluation project facilitates their seeing the interconnectiveness of a particular program in the organizational scheme. This provides a total perspective, which is in keeping with "systems thinking," a characteristic of a learning organization described by Senge (1990), as discussed in chapter 2. The process of evaluating programs lends itself well to this type of organization.

The team should represent teachers from other grades (or subjects), excluding those staff members implementing the program.

However, a limitation of a team approach for conducting evaluations could occur any time work is to be done through a committee structure. In this regard, individuals often represent diverse values and

ways of doing things. Inherent in a committee's membership is the potential for conflict and differences of opinion for accomplishing objectives. While resolution of differences can produce a better way of doing something, unless a team leader is skilled in the dynamics of group processes,[5] this agreement can detract from effective functioning.

Against this background, a principal should select evaluation team members, at least for initial projects, on the basis of how well they would work together. Unless this factor is considered, carrying out a program evaluation may be perceived as a negative experience.

FACTORS FOR PRINCIPALS TO CONSIDER IN ESTABLISHING THE EVALUATION TEAM

The prospect of having their programs evaluated is not likely to be eagerly welcomed by many faculty members. A prime reason for this is that teachers, understandably, have a great deal of ego invested in their work. It seems reasonable to assume that they believe they are doing a good job educating students. While they feel they can always improve instruction, they also maintain that they can make a difference in their students. Teachers usually know through personal observation and the "grapevine" which instructors are not very effective and fervently believe that they themselves can do a better job.

Given the egocentric[6] nature of teaching, having programs come under scrutiny could be threatening. Thus it is difficult to separate the quality of a curriculum from those who deliver it. From their perspective, if plans are made to assess their programs, the faculty may perceive that others will associate possible unsatisfactory results with the efforts of those who are responsible for implementing the curriculum. This could generate apprehension. Furthermore, teachers may resist evaluation if they believe that the students' academic levels, attitudes, and behaviors, as well as other conditions of teaching/learning, are not conducive to producing successful outcomes.

It is difficult to separate the quality of a curriculum from those who deliver it.

Faculty may also resist program evaluation if it is regarded as an additional responsibility that would take more time from their busy schedules and add another task to their already heavy workload. This perception is compounded for those staff members who feel stressed and physically tired from what they believe is a difficult teaching situation.

Another obstacle is that teachers are familiar with standardized test results and may not see any further need for determining the worth of a program because they believe that test scores are enough of a measure. Moreover, since teachers are not likely to have had experience with program evaluation, they may not understand its potential for improving educational output.

Any kind of organizational change can be difficult. Staff members are comfortable with routine, and inertia is easier than having to demonstrate new competencies called for by program evaluation. Thus, the potential for less than exuberant acceptance of participating in evaluation projects highlights the onus on leadership to promote the value of program assessment to teachers and support their efforts in conducting them. How principals can meet this challenge is discussed next.

FORMING AN EVALUATION TEAM

This section presents guidelines that principals can apply when assembling an evaluation team. They reflect various aspects of the process, including promoting, planning, implementing, and supporting the team structure as a regular part of the organization. The guidelines can be communicated by memo and/or at a meeting. Figure 5.1 lists the activities involved in the process; a discussion of each follows.

Communicate the rationale and importance of the evaluation team. Discuss the rationale for having an evaluation team as an integral and ongoing part of a learning organization. Also explain how the team can disclose information that a) confirms effective instruction and good student progress, and b) provides direction for improving aspects of teaching and learning.

FORMING AN EVALUATION TEAM: A GUIDE FOR PRINCIPALS

- Communicate the rationale and importance of the evaluation team.
- Emphasize the confirmative and diagnostic functions of evaluation.
- Solicit nominations and volunteers.
- Determine and communicate team size and tenure.
- Announce that a report on the evaluation will be made available.
- Indicate team goals and activities.
- Prepare for questions.
- Emphasize the need for collaboration.
- Describe the responsibilities of the team.
- Describe the qualifications and responsibilities of the team leader.

Figure 5.1

Emphasize the confirmative and diagnostic functions of evaluation. Convey that the evaluation team is not a "watchdog" committee whose purpose it is to find deficiencies in a program and its instruction. In contrast, emphasize the confirmatory and diagnostic functions of evaluation.

Solicit nominations and volunteers.
Encourage faculty to nominate teachers representing various grade levels and/or subjects to serve on the team and also ask for volunteers.[7] Indicate that experience in evaluating programs is a desired, but not a required, qualification for team membership.

Determine and communicate team size and tenure.
The following points should be covered:

- The team will have five to ten members (depending on the size of the school).

- The entire team is composed of faculty members who serve for an academic year with sufficient released time to carry out evaluation activities. Consider the funds available to support released time in deciding on the size of the team.
- If the number of nominees and volunteers exceed the maximum, hold an election. If there are not enough to form a team, a representative group of faculty members should then be asked to serve.
- Secretarial help will be available for typing the report and other material.
- After the year, the team should be composed of new members, except for the previous leader who would now be a member of the new team.
- By continuing in a new capacity, the former leader can share what actions were most effective and efficient as well as the pitfalls to be avoided, thus facilitating the evaluation process.

Announce that a report will be made available.
State that a report indicating the strengths of a program as well as areas that need improvement will be disseminated.

Indicate team goals and activities.
A plan for the team's goals and activities should be distributed prior to a meeting so that the staff has the opportunity to discuss its features. Ask the faculty to read it and bring their questions to the meeting.

Prepare for questions.
Anticipate the kinds of questions that might be raised. Consider the potential threat evaluating a program poses as well as other obstacles discussed previously, and think about ways to allay faculty concerns.

Emphasize the need for collaboration.
To ensure that the evaluation team plan is well-received, highlight how the team, in the true spirit of collaboration and professionalism,

can be beneficial to faculty members delivering a particular curriculum and, ultimately, to their students' performance.

Describe the responsibilities of the team.
Team responsibilities include:

- meeting with stakeholders (e.g., faculty, administrators, and parents)
- collecting and analyzing data obtained from various sources, e.g., students, parents, and those responsible for administering and implementing the program (Methods for obtaining data involve, for example, surveys, interviews, and collecting student data information.)
- forming a judgment concerning the effectiveness of the program and offering recommendations for its improvement
- writing a report describing the evaluation process and conclusions

Describe the qualifications and responsibilities of the team leader.
The team leader should be

- an experienced teacher with leadership skills.
- willing to be accountable for the team's progress.
- skilled at conducting meetings.
- task oriented.
- well organized.
- able to mediate conflict and differences of opinion.
- respected by colleagues.

THE TEAM IN OPERATION

The following guidelines are recommended for helping the team function effectively and efficiently.

The Administrator's Role

At the first meeting with the group, the principal should

- reaffirm his or her commitment to program evaluation.
- stress the importance of the evaluation team.
- remind its members that since evaluation could be perceived as threatening, they should be sensitive to this and allay any possible concerns of program staff.
- state that he or she is available for consultation and will provide any support that can be given.
- ask the team leader to serve as a liaison between the principal and the team.
- determine a date that would be best for all to hold the first team meeting.
- indicate that at this meeting a leader will be elected and that all members should review the qualifications for this role.

Preliminary Activities of the Team

At the subsequent meeting of the team, which would not be attended by the principal, the first order of business is to elect a leader and a recorder of the team's activities. The latter position is very important since it requires not only summarizing what transpired at the meetings, but also maintaining a log chronicling the evaluation activities along with completing forms to facilitate carrying them out. The log will be useful in writing the evaluation report.

After electing both individuals, dates for the next meetings should be scheduled. The leader should inform the group that, if necessary, additional meetings may be called. However, these would be held only if needed. The leader should then thank his or her colleagues for their willingness to serve on the team and state that he or she is looking forward to working with them.

Suggestions for the Team Leader

Although Schmoker (1999) advocates effective team work for school improvement, he also points out from the works of Little (1990) and

Katzenbach and Smith (1993), that meetings can be unproductive and unrewarding. According to Katzenbach and Smith (1993), unproductive discussions at meetings can result from lack of disciplined action.

In light of the potential for meetings to become drudgery, "deadly dull," or indistinguishable from others mired in routine, the guidelines shown in Figure 5.2 are designed to help team leaders take proactive steps to help ensure that meetings are events that people look forward to attending. Thus meetings should function as a tropism whereby individuals anticipate productive interaction with their colleagues and actually do gain a sense of professional and personal satisfaction in the process. Ideal as this prospect sounds, the alternative of not actively seeking this goal would likely make meetings of an evaluation team be perceived as "just more work."

Set the tone for the meetings.

Announce to the group that it is important to feel a sense of accomplishment at the end of each session. For this purpose, allow time near the close of the meeting to summarize what was achieved, and indicate the tasks that need to be completed before the team meets again. The Evaluation Plan and/or Time Log found in chapter 7 should help the team stay on track.

CONDUCTING MEETINGS—
GUIDELINES FOR THE EVALUATION TEAM LEADER

- Set the tone for the meetings.
- Ensure that each team member has an opportunity to lend input.
- Send out a meeting agenda.
- Start meetings promptly.
- Be enthusiastic.

Figure 5.2

To facilitate the work of the team, the leader should ask for volunteers to carry out various tasks. If no one offers, the leader will need to assign the particular task. A date for completion should be agreed upon. To minimize the potential for miscommunication, the leader should summarize what needs to be done for each task, unless the activity is obvious. The leader should also write down the person(s) responsible and the due date for each task so that the team member(s) can provide a progress report at a subsequent meeting. This helps ensure that aspects of the project do not "slip through the cracks." Finally, to enhance the continuity and connectiveness of the team's work, state what needs to be covered at the next session.

Meetings should function as a tropism whereby individuals anticipate productive interaction with their colleagues and actually do gain a sense of professional and personal satisfaction in the process.

Ensure that each member has an opportunity to lend input.
Explain to the team that its collective knowledge and creativity are best realized when the group engages in the activities of an evaluation project in which all members contribute. Therefore, one or two members should not dominate discussions, nor should anyone be interrupted.

Send out a meeting agenda.
The agenda should contain a reasonable number of tasks to complete in the time allocated for each session. At the first meeting, distribute an outline of the program assessment process as delineated in chapters 7–10. The team leader should be prepared to explain the various tasks involved in carrying out an evaluation study as shown in Figure 5.3.

Start meetings promptly.
Allow sufficient time, however, for people to get to the meeting place from wherever they are in the building.

STEPS IN THE EVALUATION PROCESS

- Write a program description.
- Provide direction for the evaluation.
- Formulate evaluation questions.
- Meet with stakeholders to finalize these questions.
- Obtain information to answer the evaluation questions.
- Analyze data to assess the program's impact.
- Evaluate the program's effectiveness and offer recommendations for its future development.
- Write the evaluation report.
- Assess the evaluation project.

Figure 5.3

Be enthusiastic.
Welcome the members to the first meeting. Also, maintain enthusiasm for the evaluation project throughout the entire process.

A CONCLUDING NOTE

As with any project, paying attention to details of an evaluation study increases the potential for its success. When things "slip through the cracks," they detract from the momentum of an endeavor. Time and effort are thereby wasted on trying to fix or complete aspects of an activity that should have been provided for in the first place. This is enervating and frustrating for both the leader and staff. In contrast, handling details leads to a sense of thoroughness on the leader's part, which also communicates the importance of the project to the staff.

Against this background, the chapter has presented systematic guidelines for the principal and team leader of the evaluation team to follow in the initial phases of the team's operation. For both leaders, these guidelines provide a structure to organize their respective actions, which should therefore facilitate their efforts. When thorough plans for starting and maintaining the work of the team are implemented, these increase the likelihood of enhancing the group's effectiveness and efficiency.

CHAPTER FIVE HIGHLIGHTS

1. There are several advantages of conducting a program evaluation through a team approach:
   - Opinions about the effectiveness of a program as well as its future directions can be enhanced by capitalizing on the collective professional knowledge and insights of the team members.
   - Conducting a sound evaluation requires diverse competencies.
   - Varied perspectives should result in a fair assessment.
   - A team approach makes an evaluation project more manageable.
   - Team members representing teachers from other grades and/or subjects, excluding staff members implementing the program, lend credibility to the evaluation report.
   - Having faculty from different grade levels (or subjects) serve on the evaluation team facilitates their seeing the interconnectiveness of a particular program in the organizational scheme.
2. Inherent in a committee's membership is the potential for conflict and differences of opinion. Thus, the principal should select team members, at least for initial projects, who work well together.
3. There are several factors for principals to consider in forming an evaluation team:
   - Since teachers have a great deal of ego invested in their work, the prospect of having their programs evaluated is not likely to be eagerly welcomed.
   - Having programs come under scrutiny could be threatening.
   - Faculty may regard program evaluation as an additional responsibility that would take more time from their busy schedules.
   - Teachers may not see any further need for determining the worth of a program since standardized test results are enough of a measure.
   - Teachers lacking experience with program evaluation may not understand its potential for improving educational output.

BEST COPY AVAILABLE

- Inertia is easier than having to demonstrate new competencies called for by program evaluation.

4. Guidelines that principals can apply in assembling an evaluation team cover various aspects of the process including: promoting, planning, implementing, and supporting the team structure as a regular part of the organization.

5. There are guidelines that the principal and evaluation team leader can follow to help the team function more efficiently and effectively.

ENDNOTES

1. Although not in the context of evaluating programs, the value of a team approach is supported by the research of Katzenbach and Smith (1993) as cited in Schmoker (1999). They point out that the multiple capabilities of a team are advantageous in dealing with difficult issues.

2. That a team can bring different perspectives in conducting a program evaluation is indicated by the Joint Committee on Standards (1994).

3. A program evaluation should meet the standard of being fair and complete in its assessment by identifying strengths as well as enabling weaknesses to be corrected (Joint Committee on Standards 1994).

4. The setting and conditions in which a program operates (i.e., its context) should be considered in its evaluation (Joint Committee on Standards 1994). This point is treated in chapter 8 when areas to cover in drawing conclusions about a program's effectiveness are discussed.

5. Gredler (1996) points out the importance of a moderator having this skill in conducting focus group interviews.

6. Wiggins (1996) believes that "teaching . . . is an egocentric profession in [that] . . . we find it difficult to see when our teaching isn't clear or adequate" (p. 5). However, egocentrism can also have another connotation, one that applies to teachers as well as principals. This involves the conviction that one can be influential in the change process—a belief that underscores the leadership thrust of this book.

7. Initially forming an evaluation committee through volunteers or by appointment is also recommended by DeRoche (1987). He also indicates that eventually committee membership would be elected, appointed, or formed by a combination of these methods.

8. A year is recommended to provide more opportunity for other faculty members to gain experience in evaluating programs.

BEST COPY AVAILABLE

CHAPTER SIX

MEASURING PROGRAM OUTCOMES

OVERVIEW

This chapter describes how data collected in various forms are used to measure the impact of programs. The material presented serves as a "miniguide" and as such, is not treated in depth. However, there is sufficient background to facilitate understanding of various dimensions of measurement as these relate to program evaluation.

QUANTITATIVE AND QUALITATIVE DATA

Data can be categorized into two types, quantitative and qualitative. Using numbers for recording and reporting information involves **quantitative** data. When these data are summarized in some form (e.g., means or percentages), they can be analyzed so that conclusions can be drawn about the worth of a program. Quantitative data are not limited to achievement test scores. They can reflect other outcome variables such as attitudes toward a subject or self-ratings of competencies, perceptions of a program's worth, or behavior.

BEST COPY AVAILABLE

The other type of data, **qualitative**, takes the form of verbal information, for example, responses to questions posed during interviews and/or found on survey instruments. Qualitative data provide in-depth descriptions of how situations are perceived. In this way, such data enhance understanding and insights regarding what individuals consider in assessing student growth and the rationale supporting their recommendations for program improvement.

Quantitative data, while less labor intensive to collect and analyze, are no less valuable than qualitative information. Although each type is different, they complement each other (Weiss 1998; Worthen et al. 1997), as discussed next.

COMPREHENSIVE EVALUATION

One of the criteria for sound program assessment is that the evaluation should be comprehensive in scope (Joint Committee on Standards 1994). Essentially, this means that the worth of a program should not be determined solely by one indicator, such as a group's performance on a standardized test (Joint Committee on Standards 1994). The reason for this is that the goals of a curriculum encompass more than what can be measured by a paper-and-pencil test, whether in multiple-choice or essay format.

Obviously, certain outcomes, though worthy in their own right, cannot be assessed by achievement tests—standardized or otherwise. Examples of such outcomes include the success of community service projects, attitudes toward a subject, the quality of student interaction with teachers and other students, and the confidence level shown when displaying knowledge in classroom situations.

Data from these types of indicators can be measured by colleagues observing a teacher's lesson; others can be tapped by having students complete self-rating scales. However, outcomes can also be assessed by eliciting opinions expressed in the form of qualitative data, which may corroborate measures obtained quantitatively. While judgments stated in qualitative terms are more subjective and impressionistic

because they are not obtained directly from respondents, as in the case of an attitude instrument,[1] opinions are nonetheless important evidence to have. They contribute to the comprehensiveness of the evaluation and also provide a more "human touch" to assessment information, e.g., asking a teacher *why* a particular program has helped her students improve. By obtaining illustrations to support this opinion, evidence becomes more concrete. Thus readers can relate to and/or understand information more readily, thereby extending the meaning of results reported in numerical terms.

Teacher opinions also add an important dimension to evaluation data because they are based on professional judgment. In this regard, because teachers assess their students formally and informally, they are in the best position to ascertain if any improvement in achievement, attitude, or behavior can be linked to the learning activities of a program being examined. They can make this judgment because it is grounded in their experience with other curricula and students similar to those who are (or who participated) in the current program.

Teacher opinions also add an important dimension to evaluation data because they are based on professional judgment.

Obtaining opinions is not limited to faculty. Those given by others—e.g., parents, the administrator(s) or the coordinator, as well as students—provide additional perspectives and therefore broaden the base of evaluation data. In sum, the type of qualitative data described here provides one of the valid determinants of program effectiveness.

VALIDITY AND RELIABILITY OF EVALUATION DATA

This section describes two criteria to apply in judging the accuracy of information obtained in conducting an evaluation—validity and reliability.[2] Meeting the **validity** criterion involves determining whether the information disclosed by the data collection procedure contributes relevant evidence regarding the outcome assessed.[3] To

ascertain the **reliability** of the procedure, the information it provides is examined for its consistency. Each of these criteria is discussed next.

VALIDITY

Collecting Valid Data—Some Issues to Consider

Obtaining valid evaluation data is crucial to the perceived worth of the evaluation report. This is particularly true for those delivering and administering or coordinating the program. They need to feel that the judgment of the team was based on solid evidence and that its report reflects a fair assessment. This belief makes it more likely that the report will be accepted and its recommendations put into practice.

It is not easy to accept criticism about one's work, even if it is couched in the form of positive recommendations. When teachers believe that they have provided good instruction but learn that certain changes are called for in the program, it is difficult to separate their own involvement with aspects of the curriculum and related processes that need remediation. The possibility of defensiveness, coupled with the feeling that routine is often more comfortable than change, can make recommendations for program improvement difficult to accept. This is especially so when an evaluation is based on information that individuals most directly affected by an assessment do not find creditable. This underscores the need for obtaining evaluation data believed to be valid by those whose reputation and work are affected by an assessment of their programs.

When teachers believe that they have provided good instruction but learn that certain changes are called for in the program, it is difficult to separate their own involvement with aspects of the curriculum and related processes that need remediation.

Consequently, leaders should be sensitive to the personal or human side of program evaluation, which extends beyond the technical dimension.[4] If faculty concerns are neglected or given short shrift, this could generate

resistance to the entire evaluation project and minimize cooperation—an indispensable factor in conducting an assessment inquiry. The actions discussed in the next section should help avoid the negative fallout associated with data considered not to be fair, useful, or accurate.

Collecting Valid Data—The Process

The evaluation team should invite staff teaching in the program to a meeting to elicit their input concerning the types of assessment data needed and the plan for obtaining this information. If the program has a coordinator, this person should, of course, also be involved in the data-planning phase of the evaluation project. However, it is not necessary for the principal to be included in this aspect of the process. It is not his or her responsibility to micromanage the project. Nonetheless, the principal should be kept abreast of the team's plans in progress, be alerted to any issues and problems that could have organizational implications, and be informed about the need for resources. To depend on the principal to make all decisions, especially those concerning how the project should be carried out, lessens the authority and influence of the team leader and also inhibits the professional growth of the team. Thus, the leader, in consultation with the team members, should use his or her discretion regarding the type of input sought from the principal, as well as when administrative approval is necessary for a contemplated action by the team. Giving decision-making responsibility to the evaluators should not be construed as diluting the principal's authority. Ultimately, as the instructional leader accountable for schoolwide programs, the principal has the final decision (if necessary) over those made by the team. The scope of its authority should be stated by the principal when he or she meets with the group at the first orientation meeting. Not informing the

To depend on the principal to make all decisions, especially those concerning how the project should be carried out, lessens the authority and influence of the team leader and also inhibits the professional growth of the team.

team members could lead to resentment if they had the impression that decisions regarding the project were theirs to make and not the province of the administrator.

The faculty teaching in the program should be involved with team deliberations regarding the type of data to be collected since the conclusions and recommendations of the team can affect the way teachers conduct their classes in the future. Thus the team has an ethical obligation to seek the professional perspectives and ideas of those delivering the program. Not only is it the right thing to do, but it is politically expedient because it would likely lessen objections or resistance to decisions thought to be made unilaterally by a group perceived to be too powerful. Furthermore, the input from program faculty could be valuable in its own right and facilitate decision making by the team. While other items needing remediation may come up, discussions between the team and program faculty should grow out of several considerations as indicated in Figure 6.1.

DATA COLLECTION ISSUES TO DISCUSS BETWEEN THE EVALUATION TEAM AND PROGRAM FACULTY

- Determine if the latest standardized test results are recent enough to be included as part of the pool of evaluation data. If they are outdated, consider administering a standardized test that is aligned with one or more goals of the program.
- Similarly, the results of textbook tests may be available (or could be given) if the data from these would meet the alignment criterion stated above. This information should also be used as evidence regarding the program's effectiveness.
- If the faculty of the program believes that what they have taught can best be assessed by a teacher-made instrument, then its use should be encouraged. For this, a summative type test should be developed, if one is not already available. This instrument should contain a representative sample of items that reflect student knowledge, understanding, and abilities and serve as one of the overall measures of their growth over a certain period of time.
- If the aim was to develop higher cognitive abilities and/or creativity in the form of, for example, written expression or problem solving, these types of outcomes should be tapped.

Figure 6.1

Judging the Relevancy of Tests

As with any data collection procedure, a test should first be assessed to determine if the data it provides would be a suitable indicator of a program's outcome or operation. One approach for making this assessment is to judge the content of the test according to whether its items include a representative sample of tasks that require students to demonstrate what they have learned consistent with the goals of the program. If the test meets this criterion, it can be characterized as having **content validity** since it yields information that allows one to draw reasonable conclusions about the program's effectiveness. Guidelines for judging the content validity of a test by those teaching the program are given in Figure 6.2. If more than one teacher is involved, then consensus on each of these issues must be reached.

Note that the guidelines can also enhance future performance on mandated, standardized tests since the questions raised call attention to what actions are needed to prepare students for these tests.

Responding to the questions posed in the guidelines for determining content validity of nonstandardized tests can be facilitated by using the form shown in Figure 6.3. The form is divided into two

---

JUDGING CONTENT VALIDITY OF A TEST—GUIDELINES FOR TEACHERS

- Based on prior performance on periodic tests given during the year, do most students have the knowledge and skills to answer the item correctly?
- Do students have adequate experience with the type of task posed by the item?
- Have teachers covered the material sufficiently for most students to be able to answer the item correctly?
- Overall, do the items reflect the degree of emphasis they received during instruction? For example, if teaching focused equally on math computation and solving word problems, then the test should be divided about equally between each of these competencies.

Figure 6.2

DETERMINING TEST VALIDITY

Part A
Individual items

Item Number	Material covered sufficiently (✔ if "yes")	Most students should be able to answer this item correctly (✔ if "yes")	Students have had experience with this type of task (✔ if "yes")	Decision about item: Retain Modify Reject

Part B
Entire test

Knowledge or skill area	Estimated percent of instructional time	Percentage of items covered on the test

Figure 6.3

parts. The first involves assessing each item. For the second part, a comparison is made between the estimated percentage of instructional time allocated to a particular knowledge or skill area and the percentage of items on the test that cover that area. Based on this comparison, adjustments in the number of items that relate to a topic can be made accordingly.

A Further Note on Test Validity

Judging a test for content validity enhances the accuracy of the evaluation since an instrument is used that has taken into account factors that can impact performance. Thus a fairer assessment of the program's impact can be made because students have a better opportunity to demonstrate their achievement. Moreover, there should be less frustration and better morale on the part of students as well as teachers since there is likely to be more success with tests that yield valid data.

RELIABILITY

Earlier it was stated that reliability of evaluation information refers to its consistency. **Reliability** is an essential criterion when considering the use of a data collection procedure since evidence must be dependable if it is to be accurate.

Judging Test Reliability

One way of determining whether a test provides consistent results is to check, on an item-by-item basis, what percentage of students in the top quarter of the group that took the test answered each question *incorrectly*. If this proportion is relatively large (i.e., over one-third), then it detracts from consistency of performance. When this occurs for more than 10 percent of the items, (although, admittedly, this is an arbitrary margin of error), reliability is questionable.

If an item meets the "over one-third of the group" criterion, then it should be considered a candidate for modification before inclusion in another test. Two parts of the item—the problem and the answers—should be checked. For multiple choice instruments, keep a tally of the option that was most often chosen incorrectly. If a pattern is clear, then this option should be compared to the right answer and reworded accordingly. When one wrong answer is not predominant, the problem portion of the question may need to be clarified. Teachers

should review this situation to determine whether the item ought to be revised or discarded. If any part of the item is modified, the revised version can be stored as part of a test bank for future use.

A Further Note on Reliability

Determining if information is reliable is not limited to one procedure. In this regard, there should be consistency among the types of evidence analyzed that were collected by various methods (e.g., observations, interviews, or questionnaires) in order to draw conclusions about the impact of a program. However, it should also be pointed out that, for certain outcomes, a trend in responses indicating overall support of the program among different groups may not be present. Of course, differences of opinion could occur. Where these differences are found, they do not suggest that a particular tool for collecting evidence has yielded unreliable data.

PERSPECTIVES ON AUTHENTIC ASSESSMENT

In summarizing points made by several writers, Gredler (1996) states three expectations for authentic assessments: "(1)...They should guide reform by focusing classroom effort toward more intellectual tasks, (2) . . . be more meaningful to students because they indicate real-world challenges, and (3) . . . influence the professional development of teachers" (p. 134).

Authentic assessment typically involves performance assessment and portfolios (Gredler 1996). The former, according to Haertel (1992), (as cited in Gredler 1996, p. 129) "requires examinees to demonstrate their capabilities by creating some product or engaging in some activity." Furthermore, the behavior or product assessed "is valued in its own right outside of the test setting" (Gredler p. 129). Examples of learning activities involving performance assessment include experiments, essays, and research projects (Gredler 1996).

In discussing portfolios, Gredler (1996) points out that they illustrate longitudinal growth and indicate breadth through "samples of

different kinds of work" (p. 148). Moreover, based on Wolf's (1989) conception, the ideal portfolio enables students to reflect on their own work and "become thoughtful evaluators of their own learning" (Gredler 1996, p. 154).

FOCUSING ON MEANINGFUL LEARNING AND CREATIVITY TO RAISE TEST PERFORMANCE

If we consider that the quality of our own lives is often in direct proportion to how much meaning we obtain from performing tasks inside and outside of the work setting, then students should be afforded the same opportunity to experience meaning from their work. The need for students to find learning meaningful has been identified by various writers (Beane 1995; Sergiovanni 1984; Wiggins 1989). Challenging learning activities can be meaningful if they are intriguing to the examinee, stimulate thought, and have real-world applications. These characteristics are consistent with the intent of authentic assessment. However, there is a caveat here. Unless learners are sufficiently prepared so that they can be successful at performing authentic tasks, these tasks will obviously not be very meaningful for them.

Authentic assessment is a rebellion against mindless testing that categorizes students based on efforts to achieve misdirected goals.

Portfolios provide a structure for displaying creative work. To optimize producing what is imaginative for students, the prompts or tasks that they are given should be designed to elicit creative efforts. For this purpose, teachers need to reach consensus on what type of assignments and accompanying directions are likely to result in creative productions. Such agreement increases the probability that this goal will be accomplished.

Focusing on what outcomes are meaningful (i.e., significant, consequential, and important) requires that educators determine how these outcomes should be assessed. In this regard, authentic

assessment is a rebellion against mindless testing that categorizes students based on efforts to achieve misdirected goals. The alternative, as advocated in the works of writers such as Wiggins (1989) and Wolf, LeMahieu, and Eresh (1992) that Gredler (1996) cites, is to have students challenged by genuine intellectual tasks rather than to have them learn information by rote. This lessens, and perhaps even eliminates, the resistance and ennui that often accompany tasks perceived as meaningless.

For learning to be meaningful, it should engage the student, be interesting, and even be exciting at times. This cannot happen when instruction limits multidimensional development. Humans are curious, creative, and problem solvers by nature. Instructional activities should be designed to capitalize and nurture these inherent dispositions.

What is tragic about not placing sufficient emphasis on creativity in schools is that it minimizes the benefits that enrich the quality of life for all members of society. Thus the greatest accomplishments in the arts, literature, sciences, and other fields reflect creative efforts because they are unique, innovative, and exceptional. When only a relatively small group of individuals make such contributions, the potential for creativity to elevate the quality of life for those affected by them is limited. It also means that the gratification of solving a difficult problem, producing a fine work, or giving an excellent performance is only experienced by a fortunate few. Even a creative action, thought, or product that is not noteworthy and memorable for many, can, however, be so for an individual, if it is unique to him or her and has personal value because it generates a sense of accomplishment.

The implications for what schools can do in fostering creativity for current and future generations are staggering. But creativity, as well as meaningful learning, cannot occur in a vacuum, as discussed in the following section.

LINKING AUTHENTIC AND TRADITIONAL ASSESSMENT

Instructional activities designed to achieve creative and meaningful student work need to be grounded in reality. Therefore, respectable performance concerning such work cannot occur without a foundation consisting of the *basics* of a particular discipline. High-caliber athletes and performers in other fields, as well as committed aspirants who are driven by the rewards, recognition, and gratification that excellence brings, know this, accept it, and execute routine drills until their actions become second nature. Along these lines, performing meaningful and challenging intellectual tasks that engage higher mental processes call upon a reservoir of knowledge and skills that enable cognitive functioning at an advanced level. Thus it is difficult, if not impossible, to analyze a problem situation and apply a potentially workable or appropriate solution without sufficient background.

Failing to place sufficient emphasis on creativity in schools minimizes the benefits that enrich the quality of life for all members of society.

Moreover, to be *creative* means *to create.* A creative work or performance is built upon the basics. Standardized and teacher-made tests are essential to the instructional process involved in preparing students with the necessary background to be not only creative but also equipped to handle the challenges of meaningful, intellectual tasks successfully. Achieving these goals can help make school a place that is special in the lives of students. Schmoker's (1999) position underscores the value of *using* standardized tests to serve as building blocks for realizing these goals. He states that these tests "can reveal both progress and areas needing improvement. Without derailing alternative efforts to gauge student performances and higher-order skills, standardized tests can help us make comparative assessments on how well our students are learning basic knowledge and skills" (p. 85).

For Schmoker (1999), these assessments, if based on disaggregating and thoughtfully analyzing results of standardized tests, would

"provide feedback and goal orientation" (p. 86). This effort fits within the scope of collecting and using formative (i.e., process) data to raise student achievement as discussed next.

Evaluating Curricular Processes

For individuals to improve in any endeavor, it is necessary to assess their overall performance. Such assessment identifies the area(s) that need to be "worked on" in order to do better the next time. Even though proud of a particular accomplishment, high achievers from various arenas intuitively grasp which aspects of what they have done still need to be strengthened. This sensitivity and focus, plus feedback from others, contributes to what makes them excel.

In this same vein, the *process* dimension of conducting an evaluation study provides information for areas that need improvement. Taking *specific* actions is crucial here since performance efforts must be efficient; this requires focus. When actions are focused, by definition, they are concentrated. In this regard, physical and mental energy are channeled into a synergistic force to facilitate the achievement of a particular goal.

This mentality is applicable to improving test performance whether on an authentic or a standardized instrument. With a focused orientation, an inquiry needs to be made regarding factors that affect how the curriculum is being delivered. The means for obtaining this type of formative information is discussed in chapter 7.

A CONCLUDING NOTE

If assessing a program consisted solely of measuring student achievement by standardized testing, there would be no need for an evaluation team. The worth of a program would be determined on the basis of students' performance as reflected in grade-equivalent scores or state-mandated criteria. However, evaluation is more than measurement. Assessing a program is a multidimensional endeavor that, in the final analysis, involves the exercise of professional judgment.

Such judgment is expressed in opinions regarding the intervention's effectiveness as well as recommendations for improving the program or for its termination. These opinions rely, in part, on the knowledge of basic measurement concepts and awareness of assessment issues covered in this chapter.

Assessing a program is a multidimensional endeavor that, in the final analysis, involves the exercise of professional judgment.

Since students' work should also be performance based, their achievement should be authentically assessed. When the evaluation team seeks information that reveals the extent to which traditional and authentic assessment support the program's effectiveness, the comprehensiveness of the evaluation is enhanced. Consequently, conclusions drawn about the curriculum are more definitive because corroborative, rather than narrowly based evidence, has been obtained.

CHAPTER SIX HIGHLIGHTS

1. Data can be categorized into two types, both of which should be collected in an evaluation study:
   - **Quantitative** uses numbers.
   - **Qualitative** involves verbal information.
2. An evaluation should be comprehensive, that is, a number of outcomes should be used to determine the worth of a program.
3. Evaluation data should meet the criteria of:
   - **validity**, which involves determining whether the information disclosed by a data collection procedure contributes relevant evidence regarding the outcome assessed.
   - **reliability**, which ascertains whether the information provided by the procedure is consistent.
4. In order to collect valid information, the evaluation team should elicit input from the staff delivering the program regarding the types of assessment data needed and the plan for obtaining these.
5. Authentic assessment involving challenging tasks can be meaningful if they are intriguing to the examinee, stimulate thought, and have real-world applications.
6. Traditional and authentic assessment can be linked. Standardized and teacher-made tests are essential to the instructional process involved in preparing students with the necessary background to be creative and equipped to handle the challenges of meaningful, intellectual tasks successfully.
7. Guidelines for judging the content validity of a test involve the items themselves as well as other factors relating to students and teachers.
8. Determining content validity enhances professional development because it is a reflective activity.

BEST COPY AVAILABLE

9. Reliability is an essential criterion in considering the use of a data collection procedure since evidence must be dependable (i.e., consistent) if it is to be accurate.

10. One way of determining whether a test provides consistent results is to check, on an item-by-item basis, what percentage of students in the top quarter of the group that took the test answered each question *incorrectly*.

11. There should be consistency among the types of evidence analyzed that was collected by various methods.

ENDNOTES

1. As used in this book, an instrument is broadly defined to include any tool employed for obtaining data, e.g., tests, questionnaires, or observation scales.

2. These criteria are two of those that comprise the Accuracy standard in *The Program Evaluation Standards* (Joint Committee on Standards 1994).

3. Based on *The Program Evaluation Standards* (Joint Committee on Standards 1994), a data collection procedure could consist of, for example, tests, questionnaires, observations, interviews, and student records.

4. In this vein, Posavac and Carey (1997) discuss the importance of the technical as well as the interpersonal skills of evaluators.

BEST COPY AVAILABLE

CHAPTER SEVEN

THE EVALUATION PROCESS: PHASES 1, 2, AND 3

OVERVIEW

This chapter provides the structure for obtaining comprehensive information in assessing the factors involved in delivering a program, as well as its outcomes. Following the systematic approach[1] as presented enhances an evaluation project's credibility. To this end, a variety of data sources are used to judge a program's effectiveness and efficiency.

An evaluation study begins by selecting an established program to assess, that is, one that has been functioning for at least a year. In this chapter the first three stages of conducting an evaluation are covered:

- Phase 1—Describing the Program
- Phase 2—Providing Direction for the Evaluation
- Phase 3—Obtaining Information to Answer Evaluation Questions

BEST COPY AVAILABLE

PHASE 1—DESCRIBING THE PROGRAM

In this phase, a description of the program to be evaluated is written. For this task, members of the evaluation team interview program staff to obtain information for writing the description. This is a crucial step because a sound evaluation depends on how much those judging the program know about its goals and operation. Furthermore, faculty in a program being assessed should believe that the evaluation team really knows what the teachers are trying to accomplish and how they are going about it. This adds to the perception of a fair evaluation. The areas to be covered for writing the description are outlined next.[2] A discussion of each follows.

- Rationale for the Program
- Program Goals and Objectives
- Program Operation
  - Selection of participants
  - Characteristics of participants
  - Instructional activities
  - Program personnel
  - Management activities
  - Material resources
  - Schedule
  - Costs

Rationale for the Program

If the program represents a special initiative (i.e., it is not part of the regular curriculum), then the need for it should be indicated. In such instances, describe the problem or situation that provided the impetus for developing the program. If it is not a special program, state that it is part of the regular course of study and specify the grade level(s) involved.

Program Goals and Objectives

These are presented in the form of a list. Note that objectives are more specific and are more easily measured than goals. An example of an objective would be:

> To raise grade equivalent reading scores at least one year for students who are one or more years below grade level.

In contrast, a goal may simply state:

> To improve reading achievement.

If objectives were not originally formulated, then these would not be included in the description. However, if the goals do not already exist in a document, then those responsible for delivering the program need to be surveyed to determine this information.

Faculty in a program being assessed should believe that the evaluation team really knows what the teachers are trying to accomplish and how they are going about it.

Program Operation

This part describes the structure of the program, other aspects that characterize it, and the resources that enable it to function. More specifically, program operation covers the following.

Selection of participants. Indicate how students were chosen and what criteria were used in this process. Also, specify the number of students the program served. If all students at a particular grade level are involved (e.g., a reading or math curriculum for grades 4, 5, and 6) this should be noted.

Characteristics of participants. In general terms, describe the nature of the students who are in the program, e.g., above average or at risk.

Instructional activities. Highlight the key features of how the program is being (or was) delivered with respect to both teaching and learning approaches.

Program personnel. Specify the number of people who are (or who were) involved in implementing the program and their respective roles, e.g., four teachers, one of whom served as the coordinator. Include any use of volunteers, e.g., parents or community members.

Management activities. Describe the role of the program coordinator or administrator. Specify his or her major responsibilities.

Material resources. Describe the use of any equipment or learning material in more detail than that indicated under the instructional activities section.

Schedule. State how long the program has been in operation, how often each session meets, and the length of each session. If the program is a regular, ongoing part of the curriculum (i.e., involving a particular subject area), this should be indicated, rather than how long it has been functioning.

Costs. Provide information regarding *total* costs including that expended for personnel, equipment, and material. If the program is part of the curriculum, only indicate any *special* costs, i.e., funds allocated for expenses beyond textbooks or learning materials that are regularly furnished to students by the school.

Based on the description, any reader should have a good idea of the program's purpose(s) and its operation. The description also provides background for developing and selecting evaluation questions. The information disclosed from answering these will be used to judge the effectiveness of the program. Thus these questions provide direction for the evaluation study (Worthen et al. 1997).

PHASE 2—PROVIDING DIRECTION FOR THE EVALUATION

As stated earlier, the evaluation procedures described in this book focuses on program processes and student outcomes. To obtain information on both of these dimensions, **evaluation questions** are used to guide subsequent phases of the project. Thus these questions help the evaluation team think about the kinds of data needed to answer them, the methods to be used for collecting and analyzing such data, and also form a basis for drawing conclusions about the program's methodology and impact. Furthermore, a set of evaluation questions serves as a framework that facilitates organizing the report of the study.

The following steps are involved in carrying out Phase 2:

- Formulate evaluation questions.
- Present the evaluation questions to stakeholders.
- Obtain consensus on the final choice of questions to be used.

Formulate evaluation questions.

Evaluation questions are designed to obtain information about the outcomes of the program and related processes used to achieve those outcomes. The questions should be stated so that they flow easily from each goal. Thus if it was the intent of the program to elevate reading achievement and also increase interest in reading, then these or similarly phrased questions can be used:

> Has the after school reading program been effective in raising reading achievement?
>
> Has interest in reading increased as a result of the program?

114

A question should also be raised concerning outcomes that may have occurred but which were not originally intended such as:

> Were there positive outcomes that occurred that were not indicated in the goal statements?

According to Weiss (1998), evaluators should be aware of unanticipated consequences of a program (i.e., results not intended) that are positive as well as negative. Outcomes that are negative can be identified by eliciting information about how the program can be improved. Opinions given in this regard can serve the dual purpose of disclosing areas of deficiency in the program's operation and also provide potential remedies for improvement.

The following can be placed under the umbrella of a process question that is apropos for the staff teaching in the program:

> Are there features of the program that were especially useful in helping students learn?

This question identifies particular strengths of instructional delivery. Such information can be used when plans are being made to modify the program, thereby helping to ensure that what has ben found to work will continue. In this regard, Weiss (1998) maintains that the quality of particular program features and activities and the way they are delivered will affect the extent to which outcomes are successful.

In addition to determining the extent to which a program has been successful in achieving its outcomes—the effectiveness dimension of evaluation—another criterion to investigate is the curriculum's efficiency. This aspect concerns judging if the outcomes produced were worth what they cost (Weiss 1998). An economic perspective on worth is appropriate here because efficiency concerns the issue of whether resources expended to operate the program were put to best use, that is, were well spent. With this in mind, the following evaluation question adds another dimension to the assessment process by tapping opinions of program staff and the administrator(s):

> Are the outcomes produced by the program worth the time and/or money expended to prepare for and implement it?

The information that would be obtained by this question is not, by itself, typically used to judge program efficiency. What is commonly carried out for this purpose is, according to Weiss (1998), a technical approach that is usually in the form of cost-benefit and cost-effectiveness analyses.[3] This situation notwithstanding, an evaluation question eliciting opinions about resources used, while subjective and not expressed in economic or accounting terms, is nonetheless useful. The reason for this is that such opinions are grounded in the professional experience of those who deliver or are responsible for administering the program. Thus if the weight of evidence indicates that resources were well spent, this reinforces the value of the program. In contrast, if information reveals that financial and/or time expenditures were too high relative to program outcomes, then such information could (a) serve as a stimulus for thinking about how resources can be better spent or (b) lead, as discussed by Weiss, to the adoption of a more promising program.

The following is a generic evaluation question about curriculum processes that should be raised with different groups:

> Do you have any recommendations for improving the program?[4]

How this question can be extended to obtain specific suggestions to strengthen the program will be described in Phase 3.

To complete the set of evaluation questions ask:

> Do you have any additional comments about the worth of the program?

Note that the language used for evaluation questions can be the same (or similar to) that used for interviews or found on questionnaires that require a written response. How these questions are to be phrased and for whom they would be appropriate depend on the group surveyed.

Since open-ended questions provide the opportunity to obtain more in-depth information, those surveyed should be asked to explain their responses.

For a site-based evaluation project, both of the following criteria[5] should be applied in formulating and selecting the final list of questions:

- How important would the information elicited by this question be to decision makers and others able to improve the program?
- Is it feasible to answer this question given the availability of (1) the information sources and (2) time?

After the program's goals and questions have been written, copies of both should be distributed to the *stakeholders*. If the group is too large (e.g., parents), the materials can be distributed to a representative sample. Consistent with Weiss (1998) and Worthen et al. (1997), to decide who qualify as stakeholders, be guided by considering which groups or individuals have a "stake," i.e., a strong enough interest in the program's success so that their input should be included. However, their inclusion should not be for political reasons only, but also to bring other useful perspectives.

Ask stakeholders if they have suggestions for other questions. Indicate that these should be formulated to evaluate the *overall effectiveness* of the program. Put another way, the questions should help determine if the program is successful thus far (or was successful). For this purpose, stakeholder questions should aim at measuring results or outcomes in light of the program's goals.

Present the evaluation questions to stakeholders.

In this step, the stakeholders and the evaluation team judge which questions should be addressed by the evaluation study based on the two previously stated criteria—their importance and feasibility. This determination should yield a more manageable set of questions.

Obtain consensus on the final choice of questions to be used.
Consensus should be based on agreement between the evaluation team and the stakeholders. For Worthen et al. (1997), agreeing on evaluation questions by consensus results in a feeling of "shared ownership" which "greatly enhances the probability that evaluation findings will be used" (p. 261).

PHASE 3—OBTAINING INFORMATION TO ANSWER EVALUATION QUESTIONS

After the evaluation questions have been agreed upon, the process of collecting evaluation information may begin.

Preliminary Considerations

Decide if a control group should be included as a basis of comparison to judge the program being assessed. If a comparable group is not being used, baseline data (e.g., pretest measures such as test results and report card grades) are needed to judge the extent of improvement.

The question of whether to use a sample or the entire population should be resolved. The following speaks to *not* using a sample:

- If only a relatively small number of individuals are involved, then using a sample is unnecessary (Weiss 1998; Worthen et al. 1997). For example, collecting information from two to four classrooms would not be too time consuming.
- Since the school has results of standardized tests, this information may be readily available if the evaluation plan calls for such information.
- If an additional non-standardized test is being administered as an outcome measure to determine program effectiveness, teachers would have the scores from this test anyway. The evaluation team could divide the task for summarizing these data (e.g., by calculating means and percentages) based on the performance of all students. The same holds true for report card information.

Using a Sample

The prime reason for sampling is that the resources available to the team in terms of personnel and time may be limited if the data from the entire population were to be collected and analyzed. Given this possibility, the team has to estimate what is feasible to accomplish. Therefore, a sample size that is manageable needs to be determined. For this purpose, the guidelines in Figure 7.1 should be helpful.*

The numbers in Figure 7.1 are based on drawing a sample from a single population. However, if program participants and nonparticipants are being compared, then appropriately sized samples should be drawn from each of these populations.

After the sample size has been determined, an expedient way of obtaining the sample is to select *classes* (rather than individuals) so that their *total* enrollment is at least that of the size desired. The process for choosing samples in this way is called **cluster sampling** where each group, in this case a class, is known as a cluster. The steps for choosing sample classes are:

1. Number each class on separate slips of paper, fold, place in a container, and mix the contents.
2. Draw one slip, note the class number, and replace the paper in the container. The reason for replacing the slip is to keep the probability of selecting a sample constant, i.e., one class out of the total number of classes.
3. If the same group is drawn again, it is ignored since it was already chosen. Replace the slip and pick another.

A Note on Selecting Classes Randomly

If students are enrolled in a special program being assessed (e.g., one offered after school) then selecting entire classes may not *initially* provide the required sample size. The reason for this is that both participants and nonparticipants may be attending the same classes in which a survey will be distributed. Therefore, it would be necessary to have a place on a questionnaire that asks students to designate whether or not they were participants. If the sample size is not

POPULATION SIZE AND REQUIRED SAMPLE SIZE

Population	Sample
300	180
400	208
500	230
600	246

Figure 7.1

obtained after the first administration of a questionnaire, then a second set of classes can be randomly selected shortly afterwards to make up for the deficit in sample sizes.

A further point: When only a sample of classes is chosen to participate, even randomly, the assessment of the program may be criticized by teachers and parents as being neither fair not accurate since input was not obtained from all classes. More specifically, consider the situation where the program was judged as only moderately successful. However, a number of teachers whose classes performed relatively well were excluded from the sample. They may feel that the report does not represent their students. With respect to parents, it is good public relations to inform them of program changes. When improvements are planned, they would like to know that their input was considered. Sending questicnnaires to the entire population provides this opportunity as well as broadening the base of support.[7]

In sum, when all members of various stakeholder groups have the chance to contribute evaluation data, this enhances ownership in the revised program. Another issue is that after doing all of the work required for an assessment project, the evaluation team should not be in the position of finding that their work is *unnecessarily* criticized. This could have an adverse effect on its morale as well as dampen the spirit of the succeeding team.

Nevertheless, there is a reality to consider when conducting focus (small) group interviews with students. If these were to involve a fairly

large group, say 200 (or even more), then such interviews may take too much time and thereby delay issuing the evaluation report. In this regard, the evaluation project should not be overly ambitious. If this occurs, the work could become onerous, rather than productive. Therefore, using a smaller group may be more feasible. How to obtain a sample for these interviews is discussed next.

Selecting Focus Groups

One approach for deciding on the sample size to use in conducting focus group interviews is to have the evaluation team decide on how many of these are practical to schedule. For example, the decision might be to hold twenty interviews, each involving a group of five students—ten group interviews for the innovative and ten group interviews for the regular program (fifty students from each set). Assume that the population of each program is approximately 150 students. Two separate lists are composed, one containing names from the innovative program and the other from the regular group. Since the sample size is one-third of 150, then every third name is selected from each list until there are 50 students representing each program.

Selecting respondents in this way is known as **systematic sampling**. While this method is not random sampling because everyone did not have an equal chance of being selected, systematic sampling should produce a representative sample more efficiently than by drawing names from a container. Systematic sampling can, of course, be used if there is no control group.

Note: While it is important to plan data collection procedures, the team may find that after it has begun conducting focus group interviews, there is time to use this method to obtain information from *all* students.

With respect to teachers, all who have taught in either the innovative or regular program should be surveyed. First, there is greater opportunity to obtain a diversity of opinions, both favorable and unfavorable. Second, it is more political to have greater inclusiveness. Also, not involving everyone could lessen the report's credibility because some teachers who were not surveyed could feel that their

opinions did not count. Moreover, having a greater number of opinions, including input from administrators and coordinators, provides the opportunity to obtain more reliable evidence because one can better determine if there is consistency among the responses.

Data Collection Methods

Methods for collecting data should be selected to address each of the evaluation questions. This requires tapping a variety of information sources using multiple procedures (e.g., test scores, questionnaires, and interviews) in order for the evaluation to be comprehensive (Joint Committee on Standards 1994). When this criteria is met, there is greater opportunity for the results to be consistent which would thereby strenghten conclusions drawn about the program.

Use **unobtrusive measures** consisting of records that relate to student outcomes such as report card grades, norm- or criterion-referenced test results, or discipline records. When individuals are being observed or know that they are involved in a study, this awareness can affect the way they act or respond. Since this situation cannot occur when data are obtained from a source such as records, this type of information can be categorized as an unobtrusive measure.

An important dimension of the evaluation is measurement of attitudes. For this outcome, use such indicators of student development as self-confidence and/or attitudes toward learning related to the program under study. Also assessed are the attitudes (i.e., opinions of the program) held by stakeholders who are knowledgeable about its goals and outcomes.

To determine student attitudes, first check for instruments that may be available in the school or district. If this does not produce any that are related to the program, then an ERIC "subject" search (including the Current Index to Journals in Education, i.e., CIJE) can be made using the following descriptors: student attitudes, subject area (e.g., mathematics), and questionnaires.

If one of the goals of the program is to develop self-confidence as it relates to a particular subject, then an ERIC "title" search can be

made using "academic self-concept" and then selecting the subject area from the collection of articles on this topic.

Using the ERIC database should disclose an instrument consistent with the goals of the program. If, however, the team in consultation with the program staff feels that developing its own questionnaire may be more suitable for the attitudes the program sought to foster, then an instrument can be developed for the evaluation study. Examples of attitude items are given later in this chapter. Keep in mind that unless otherwise indicated, permission should be sought to use or modify a copyrighted instrument.

Also useful are **interviews** that enable one to gain additional insights about the effects of the program. This is likely to occur because of the opportunity interviews provide for obtaining in-depth responses of stakeholders. However, information obtained by this method is more difficult to analyze as compared to attitude scales because the latter lend themselves more easily to quantification.

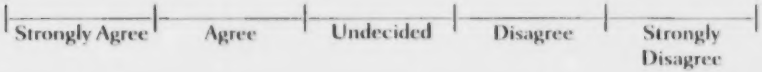
Interview questions can also be the same as (or similar to) those used for the evaluation questions. This is also the case for open-ended items on a survey questionnaire. No matter how the data collection procedure elicits information from respondents, the language should be phrased so that it is appropriate to the perspectives of different stakeholder groups.

For example, the following evaluation question can also be used to obtain information from interviews conducted individually with program staff in which the respondent is asked to support his or her answer:

> Has interest in reading increased as a result of the after school reading program? Please explain.

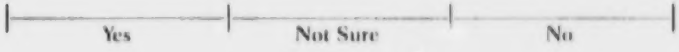
On a survey questionnaire sent to parents, the question can be phrased as a statement to record the degree of agreement with the issue raised as measured by a Likert scale. For example:

My child's interest in reading has increased because he or she has been in the after school reading program.



For a survey completed by students, the statement might read:

I like to read more because I have been in the after school reading program.



When constructing or using an instrument, the evaluation team should determine if students have the cognitive maturity to make the judgments called for on the gradations of the scale. Thus, for older students a five-option scale could be appropriate.

For interviews with students, the surveyor might ask:

Do you read more because you have been in the after school reading program?

As previously discussed, interviews can be held using small focus groups (e.g., four to five students). With parents, interviews can be done by phone if it is more expedient to do so. Selecting parents to interview can be done by the systematic sampling method described earlier. The sample size would depend on the amount of time the evaluation team can devote to this activity. As with any interview, the focus should be on eliciting in-depth information that cannot be obtained when respondents choose options on a Likert scale incorporated in a survey questionnaire.

Follow-Up Questions

If students who participated in the intervention are now in a higher grade, or if they were in an ancillary program such as one offered after school, then the current teacher(s) can be asked follow-up questions. Answers to these provide additional information on program outcomes, thereby disclosing evidence regarding how durable the effects of the intervention are with respect to students who are no longer participants. The following is an example of a follow-up question:

> Are these students, as compared to others you now have (or have had), achieving particularly well and if "yes," in which way?

Data can also be obtained from these teachers regarding other indicators of program effectiveness, e.g., interest in reading. These types of questions would be formulated based on the goals of the program. However, a generic question should be raised, such as:

> Have you observed anything special about these students concerning the way they perform or do their work as compared to others?

The answers may reveal a positive outcome that was not the intent of the program, e.g., a higher level of self-confidence.

The information obtained from each follow-up question should be subsumed as evidence under its appropriate evaluation question.

Obtaining Formative Information

Recall that one of the evaluation questions asks if there are recommendations for improving the program. This information, from the perspective of those delivering the program, can be obtained through a needs assessment as shown in Figure 7.2. The instrument is comprehensive; it covers issues of curricular content, instructional activities (processes), and learning materials. Opinions are also elicited involving a variety of factors that can affect how well the program functions and consequently, student outcomes. Since staff members

PROGRAM NEEDS ASSESSMENT

Name of Program:_____

Directions

Listed below are items regarding a variety of factors that can have an impact on the effectiveness of the program being evaluated.

- For each item, please indicate the extent of need for change by checking the appropriate box.
- Indicate any recommendation(s) that could improve delivery of the program or its operation.

	Extent of need for change:			
	None needed	Low priority	Medium priority	High priority
1. Inservices for teachers*	❑	❑	❑	❑
Recommendation(s) for improvement:				
2. Scheduling of the program	❑	❑	❑	❑
Recommendation(s) for improvement:				
3. Changing the textbook	❑	❑	❑	❑
Recommendation(s) for improvement:				
4. Modifying or adding other learning materials*	❑	❑	❑	❑
Recommendation(s) for improvement:				
5. Instructional methods	❑	❑	❑	❑
Recommendation(s) for improvement:				

*Items 1, 4, 6, 7, 10, 11, 13, 20, and 23 are based on material indicated by DeRoche (1987) concerning aspects of a curriculum.

Figure 7.2

	Extent of need for change: None needed	Low priority	Medium priority	High priority
6. Changing the curricular content*	❑	❑	❑	❑
Recommendation(s) for improvement:				
7. Changing the sequence of curricular content*	❑	❑	❑	❑
Recommendation(s) for improvement:				
8. Learning activities are motivating.	❑	❑	❑	❑
Recommendation(s) for improvement:				
9. Eligibility of students for participation in the program**	❑	❑	❑	❑
Recommendation(s) for improvement:				
10. Teachers need to meet more frequently to discuss and/or collaborate on articulation (i.e., continuity of learning activities across grade levels) and other curriculum issues as well as instruction*	❑	❑	❑	❑
Recommendation(s) for improvement:				

**This item would apply to ancillary or other programs that are distinct from the regular curriculum (e.g., an after school or honors program).

Figure 7.2 (continued)

	Extent of need for change: None needed	Low priority	Medium priority	High priority
11. Productiveness of these teacher meetings with respect to the discussions and/or collaborative efforts*	❑	❑	❑	❑
Recommendation(s) for improvement:				
12. Based on regular testing, follow-up is geared toward the specific area(s) of deficiency.	❑	❑	❑	❑
Recommendation(s) for improvement:				
13. Learning activities are suitable for students with different ability levels.*	❑	❑	❑	❑
Recommendation(s) for improvement:				
14. Students have received training on how to respond to questions posed on different types of exam formats e.g., multiple choice.	❑	❑	❑	❑
Recommendation(s) for improvement:				
15. Students understand what a question is asking and therefore know how to supply the correct answer.	❑	❑	❑	❑
Recommendation(s) for improvement:				

Figure 7.2 (continued)

	Extent of need for change: None needed	Low priority	Medium priority	High priority
16. Students realize the need to read a test item carefully in order to know what a question calls for.	❑	❑	❑	❑
Recommendation(s) for improvement:				
17. Students understand the vocabulary found in various types of questions and/or assignments that have special meaning for a particular subject.	❑	❑	❑	❑
Recommendation(s) for improvement:				
18. Students have experience in taking "practice" standardized tests.	❑	❑	❑	❑
Recommendation(s) for improvement:				
19. Students' explanations regarding how they solved the problem or why they gave an answer to a question has provided insight to guide subsequent instruction.	❑	❑	❑	❑
Recommendation(s) for improvement:				
20. Students have a solid foundation in the basics of a subject.*	❑	❑	❑	❑
Recommendation(s) for improvement:				

Figure 7.2 (continued)

	Extent of need for change: None needed	Low priority	Medium priority	High priority
21. Students have sufficient knowledge, understanding, and skills to learn more advanced subject matter.	❑	❑	❑	❑
Recommendation(s) for improvement:				
22. Students complete all or most homework assignments at a satisfactory level.	❑	❑	❑	❑
Recommendation(s) for improvement:				
23. Students have experienced being successful on tests and assignments on a regular basis.*	❑	❑	❑	❑
Recommendation(s) for improvement:				
24. Students feel confident that they can obtain good or even excellent test scores and grades.	❑	❑	❑	❑
Recommendation(s) for improvement:				
25. Students are able to learn more challenging material.	❑	❑	❑	❑
Recommendation(s) for improvement:				

Figure 7.2 (continued)

	Extent of need for change: None needed	Low priority	Medium priority	High priority
26. Students' attitudes are positive.	❑	❑	❑	❑
Recommendation(s) for improvement:				
27. Efforts are directed at soliciting support and involvement of parents with program activities.***	❑	❑	❑	❑
Recommendation(s) for improvement:				
28. Are there any other aspects of the program that need attention? Please specify level of priority.	❑	❑	❑	❑
Recommendation(s) for improvement:				

***This item is based on an indicator of effective schools reported in reviews by Purkey and Smith (1983 and 1985) as cited in DeRoche (1987).

Figure 7.2 (continued)

consider each item, the questionnaire provides a standard frame of reference. Thus the formative information yielded by the instrument should result in more efficient efforts directed at improving the program because time, energy, and resources can be concentrated on areas of greater need.

The questionnaire should be completed anonymously by the program staff. With respect to improving an aspect of the program's operation, the respondents are asked to indicate the extent of need for change. Also requested are any recommendations for improvement. After the results are tabulated, focus group interviews with the program staff can be scheduled concerning recommendations that require elaboration or further clarification. The interviews would especially target those suggestions where a trend in the data is evident for "medium" and "high" priority items.

Focus group or individual interviews with program staff can also be conducted to address the other evaluation questions. The decision to use either method or a combination of these for any of the questions depends on the time available to the team. It may be necessary to schedule more than one session. It would be helpful to explain to the program staff that an interview format usually provides more in-depth opinion than would be obtained from a questionnaire alone.

Conducting Observations

Another evaluation method involves classroom observations. The procedures recommended for conducting these are given in Figure 7.3.

Carrying out observations can provide useful information since evaluators then have another source of tangible evidence concerning the program's effectiveness. However, teachers may resist the prospect of peers coming into their classrooms to assess the extent of learning taking place. Thus they may not feel comfortable with observations because these can easily be construed as "Not only is my program being evaluated, but so is my teaching." Consequently, permission for having the evaluation team members visit classrooms should be obtained before any observations are made.

PROCEDURES FOR CONDUCTING CLASSROOM OBSERVATIONS

Conduct classroom observations at consistent times.

Visits should take place during the same time period (i.e., a session in the morning or afternoon) and within a certain time frame (e.g., two weeks).

Use more than one observer.

This lends reliability to the measurement.

Carry out a series of observations.

Sufficient data are needed in order to draw more reliable conclusions about classroom activities. If only one observation is made, a class may have had either an "off" day or a particularly productive one. Thus it is recommended that two to three observations be conducted.

Observe a control group.

If available, a control group should also be observed to provide a basis of comparison.

Figure 7.3

Given the possibility of a negative reaction to observations, the principal's leadership is important in conveying to the teachers that they are not being evaluated as *individuals*. The principal should also communicate that all professionals can improve through feedback, and that the delivery of a program can be enhanced in this way as well. This position is consistent with the thrust of a learning organization. In efforts to allay possible concerns, teachers should be reminded that sound program evaluation not only confirms learning activities that are effective but also diagnoses areas that can be improved.

A rating scale such as the one shown in Figure 7.4 can be used to record observations. The list of items is not meant to be exhaustive and can be modified to suit specific situations. Moreover, this type of instrument can serve the dual role of eliciting program processes and outcome data. Comments given in the space provided below each item can provide a source of information for making recommendations to improve the program.

CLASSROOM RATING SCALE

Date:_____

Observer:_____

Goal(s) of the lesson observed:

1. _____
2. _____

Students are participants in the program being evaluated.* ____Yes ____No

1. Students were actively engaged in the learning process.
   1 2 3 4 5 NA (not applicable)
   Low High
   Comment: _____
   _____
2. Students responded well to the teacher's questions.
   1 2 3 4 5 NA
   Low High
   Comment: _____
   _____
3. Students raised "how" and "why" questions.
   1 2 3 4 5 NA
   Low High
   Comment: _____
   _____
4. Students were attentive when the teacher gave explanations.
   1 2 3 4 5 NA
   Low High
   Comment: _____
   _____

*This information is included when an innovative program is being compared to a regular one.

Figure 7.4

5. Students appeared well prepared for the lesson.

1	2	3	4	5	NA
Low				High	

Comment: _____

_____

6. Students displayed appropriate classroom behavior.

1	2	3	4	5	NA
Low				High	

Comment: _____

_____

7. Students appeared to be interested in the lesson.

1	2	3	4	5	NA
Low				High	

Comment: _____

_____

8. Students used their time productively.

1	2	3	4	5	NA
Low				High	

Comment: _____

_____

9. Students responding to the teacher's questions was not limited to a few individuals.

1	2	3	4	5	NA
Low				High	

Comment: _____

_____

10. Students demonstrated understanding of the lesson by their responses.

1	2	3	4	5	NA
Low				High	

Comment: _____

_____

Figure 7.4 (continued)

11. Students held an effective dialogue with the teacher.

1	2	3	4	5	NA
Low				High	

Comment: _____

_____

12. Students held an effective dialogue with each other.

1	2	3	4	5	NA
Low				High	

Comment: _____

_____

13. Students made appropriate references to the text.

1	2	3	4	5	NA
Low				High	

Comment: _____

_____

14. Students were able to use learning material besides the text effectively.

1	2	3	4	5	NA
Low				High	

Comment: _____

_____

Figure 7.4 (continued)

Organizing Procedures for the Evaluation Project

Before collecting any type of data, plan for this activity by completing the first six categories in the Evaluation Plan, a model of which is shown in Figure 7.5. (*Note:* The space below each item can be expanded to accommodate more information if necessary.) Before completing category 7, read chapter 8, which discusses various types of data analyses. The plan (including copies of interview questions and survey questionnaires) should be presented to the principal, program staff, and other stakeholders to obtain feedback and consensus.[8]

Schedule project activities by completing column 1 of the Evaluation Time Log (Figure 7.6). As each activity is finished, enter the date in column 2. This should facilitate planning the team's work since one would be able to see the extent to which the deadlines set have been met.

Ethical Aspects of Program Evaluation

In assessing the worth of a program, the evaluation team faces several ethical issues. Among those that Weiss (1998) discusses is the need to maintain confidentiality and anonymity of all information obtained in an evaluation. Aligned with Weiss' point, when information is not identified as having come from a specific source, this has implications for building trust and cooperation not only for the current evaluation study but for future projects as well. Thus individuals should not feel threatened or reluctant to provide evaluation data for fear that their program may be assessed as less than satisfactory. Such an assessment could be perceived by the program staff as lowering their status among members of the school community. People should therefore feel free to express their opinions (Weiss 1998) in the context of providing accurate information that can improve program delivery. Consistent with Weiss, such freedom can occur when those interviewed believe that by giving their opinions of a program, they will not be identified by name in the evaluation report, nor informally by the evaluation team to administrators and other faculty members.

EVALUATION PLAN

1. Evaluation question:

   _____

   _____

2. Information requirements (What information is needed to answer the question?):

   _____

   _____

3. Information source (Where and how can the information be obtained?):

   _____

   _____

4. Instrument(s) or measurement(s) used:

   _____

   _____

5. Design (Indicate comparisons to be made, e.g., comparing two groups on a test or pre- to posttest gains for one group.):

   _____

   _____

6. Schedule for obtaining data (Set date[s] for collecting this information.):

   _____

   _____

7. Data analysis (How will results be summarized to help answer the question?):

   _____

   _____

From *Program Evaluation Skills for Busy Administrators* by T. R. Owens. © Northwest Regional Educational Laboratory. Adapted with permission.

Figure 7.5

EVALUATION TIME LOG

	1 Completion Date Set	2 Date Activity Completed
1. Determine the scope of the evaluation.		
a. Write the program description.	_____	_____
b. Formulate the first set of evaluation questions.	_____	_____
c. Agree with stakeholders on final set of evaluation questions.	_____	_____
2. Plan the evaluation.		
a. Complete parts 1–4 of the Evaluation Plan.	_____	_____
b. Complete part 5–7 of the plan.	_____	_____
3. Implement the evaluation.		
a. Administer instruments as scheduled in the plan.	_____	_____
b. Collect other data.	_____	_____
c. Analyze data.	_____	_____
4. Report the evaluation study.		
a. Submit draft copy to stakeholders.	_____	_____
b. Submit final report.	_____	_____

From *Program Evaluation Skills for Busy Administrators* by T. R. Owens. © Northwest Regional Educational Laboratory. Adapted with permission.

Figure 7.6

In the case where there is only one faculty member providing instruction in the program being evaluated (e.g., an honors class), then maintaining confidentiality and anonymity is, of course, not possible. However, if the evaluation report is to be disseminated beyond the district, it is best that the teacher's name not be indicated unless permission is given to do so.[9] A further point: If the environment of the school truly reflects a learning organization, then a teacher should not be concerned about expressing his or her opinions since such input could be useful in improving the program. Such freedom is an outgrowth of a principal's leadership in promoting and supporting professional and student growth through experimentation with innovative and promising curricular changes.

If the evaluation report is to be disseminated beyond the district, it is best that the teacher's name not be indicated unless permission is given to do so.

Another ethical issue that Weiss (1998) discusses is informed consent. This means that the evaluation team must respect the rights of individuals who do not wish to provide information on any aspect of the program. Along these lines, the team must adhere to the policy of the district with respect to obtaining parental permission for students to give input concerning their experiences with an innovative or regular program as well as other student data.

Guidelines for Collecting Evaluation Data

This section synthesizes the approach to data collection procedures discussed in this chapter and also covers other aspects to consider when carrying out these procedures.

1. If an evaluation study involves a comparison between an experimental and a control group, or if only a single group is being assessed, the same set of evaluation questions is used.

2. With a two-group design, separate interview questions should be formulated so that they are applicable for the experimental and control groups respectively. For example, the staff teaching in the innovative program could be asked:

How effective is the innovative program in raising math scores?

Very effective	Effective	Somewhat effective	Not very effective

For instructors in the regular program, the same type of question and scale is posed. However, the word *current* is substituted for *innovative*.

The following question is apropos to those currently teaching *both* the experimental and control classes:

As compared to the regular curriculum, how effective is the innovative program in raising math scores?

Very effective	Effective	Somewhat effective	Not very effective	Equally effective	Neither is effective

Note that the scale accompanying the interview questions should facilitate making a judgment by the respondent. The reason for this is that a continuum provides a frame of reference that clarifies the relative choices available. Another advantage of using such a scale is that it makes it easier to analyze and report responses. For example, a report might say:

Three out of four teachers in the innovative program believed it to be very effective in raising math scores. One felt it was only somewhat effective.

Whether or not a scale is used in conjunction with interview questions, respondents should be asked to explain their answers. Such explanations provide further evidence to support the opinions given.

3. When using experimental and control groups, the items on a survey questionnaire should be generic so that they apply to both programs. As previously stated, the Likert scale on which the respondent records his or her answers should be appropriate for the person's age. For example:

What I am learning in my English class helps me in my other subjects.

Strongly agree	Agree	Not sure	Disagree	Strongly disagree

It's easy for me to learn new things in math.

Yes	Not sure	No

I like the discussions we have in my social studies class.

Strongly agree	Agree	Not sure	Disagree	Strongly disagree

My science class is interesting.

All of the time	Most of the time	Sometimes	Never

Generic items on a survey questionnaire (as well as interview questions appropriate for each group) provide a current basis

of comparison. In this regard, students in the innovative program are measured against those in the regular curriculum. For a one-group design, interview questions and survey questionnaire items would, of course, refer only to the intervention. Following are examples.

Interview question for staff:

To what extent has the innovative program been effective in raising math scores?

Very effective	Effective	Somewhat effective	Not very effective

Questionnaire item for students:

I can read better because I have been in the after school reading program.

Yes	Not sure	No

Note that all of the survey questionnaire items have been stated positively. These can be phrased this way if the instrument is relatively short. However, if a longer questionnaire is constructed, then the possibility of *response set* is increased. This possibility refers to the tendency of an individual to answer a certain way, for example strongly agree (or agree) throughout the questionnaire. Interspersing some negatively stated items could lessen this tendency.

The following shows how such items would be *reverse scored*. Note that the most positive response is worth 5 points using the following Likert scale.

My science class is boring.

	Strongly agree	Agree	Not sure	Disagree	Strongly disagree
Points assigned	1	2	3	4	5

The next item elicits the same opinion, but is stated positively with the points indicated for each response.

My science class is interesting.

	Strongly agree	Agree	Not sure	Disagree	Strongly disagree
Points assigned	5	4	3	2	1

Whether the item is stated positively or negatively, the middle option on a 5-point scale is still worth 3 points.

4. Both quantitative and qualitative data should be obtained whenever it is possible to do so. Having both types serves to corroborate conclusions drawn about a program's effectiveness.

CHAPTER SEVEN HIGHLIGHTS

1. The following constitute the first three phases of program evaluation:
   - Describing the Program
   - Providing Direction for the Evaluation
   - Obtaining Information to Answer Evaluation Questions
2. Describing the program is a crucial first step since a sound evaluation depends upon how knowledgeable those judging the program are about its goals and operation.
3. Evaluation questions accepted by stakeholders through consensus are designed to obtain information on program processes and outcomes. These questions provide direction for subsequent phases of the assessment project.
4. Various data collection procedures should be used for a comprehensive evaluation.
5. An evaluation plan and an evaluation time log serve to organize the work of the team.

ENDNOTES

1. The evaluation process presented in chapters 7 and 8 is based on the approach found in the works of Owens (1979) and Worthen et al. (1997).

2. The program description headings have been modified from Owens (1979).

3. See, for example, Posavac and Carey (1997), Weiss (1998), and Worthen, et al. (1997) regarding methods of determining program costs.

4. If the program has been completed, ask how it could be improved should the decision be made to continue it in the future as in the case of submitting a proposal for re-funding.

5. These criteria have been modified from the following references respectively: Owens (1979) and Worthen et al. (1997).

BEST COPY AVAILABLE

6. The required sample sizes are somewhat higher than those indicated by Krijcie and Morgan (1970). This was done to compensate for any questionnaires that are not usable because they were completed incorrectly or have missing information. A comprehensive table of required sample sizes based on the total number in a population may be found in Krijcie and Morgan's article.

7. That stakeholders (i.e., parents) who are excluded from the sampling process may not be as supportive of a program is a real possibility. The author would like to thank Dr. Cozette Buckney, an experienced building and central office administrator, for confirming this point. Further support may be noted from *The Program Evaluation Standards* (Joint Committee on Standards 1994) which analyzes cases, most of which are based on actual situations. One case describes a scenario where the credibility of the evaluation report was challenged, in part, because the samples were believed to be unrepresentative of their populations. It was later decided to terminate the program.

8. This step is in line with the recommendations that participation by stakeholders should be fostered in an evaluation (Worthen et al. 1997) and that they should be given the opportunity to review instruments and data collection procedures (Joint Committee on Standards 1994).

9. Weiss (1998) also raises the point that individuals quoted in a report should not be identified without their permission.

CHAPTER EIGHT

THE EVALUATION PROCESS: PHASES 4, 5, AND 6

OVERVIEW

This chapter describes the later steps in completing an evaluation project with a view toward improving a program. These involve the following phases:

- Phase 4—Analyzing Data to Assess a Program's Impact
- Phase 5—Evaluating the Program's Effectiveness and Efficiency and Offering Recommendations for its Future Development
- Phase 6—Writing the Evaluation Report

PHASE 4—ANALYZING DATA TO ASSESS A PROGRAM'S IMPACT

The use of quantitative and qualitative data involves different analytical approaches. These approaches complement each other, however, when conclusions are drawn from the results disclosed by responding to the evaluation questions.

The quantitative data analyses described in this section are based on the criterion of practical significance, rather than statistical significance. Although the latter provides an objective gauge for assessing a

BEST COPY AVAILABLE

program's impact, the overriding standard to use for weighing outcomes is one of practical significance. The basic question to ask to determine such significance is: Have the students *made meaningful* progress in this program?[1] The reason for this criterion is that a relatively small gain for program participants, say three or four months, may turn out to be statistically significant (see, for example, Hopkins and Glass 1978; Vaughn 1998; Worthen et al. 1997). This indicates that the difference would be due to chance rather than the intervention a very small portion of the time (e.g., 5 times in 100). However, in view of the amount of time and resources expended in implementing instructional activities, a small gain, although significant, may not be considered that important. Thus the final determination of how successful a program is rests with the evaluators. Their judgment as practitioners is grounded in the collective knowledge of the team members whose experience in formally and informally assessing student achievement, attitudes, and behavior has prepared them for participation in an evaluation project.

The final determination of how successful a program is rests with the evaluators.

Relying on practical significance as the fundamental standard for evaluating outcomes does not preclude conducting statistical analyses to substantiate findings. In this regard team members knowledgeable about how to analyze computer-generated results can use a computer program to determine if a difference is significant. However, a statistical test that can be readily applied without the use of a computer is described in the next section. In any case, the team's objective is to produce a credible evaluation that reflects a systematic and comprehensive effort to determine program effectiveness.

A Statistical Method for Comparing Groups

A test of significance (e.g., a t-test) discloses the extent to which chance is responsible for causing the difference between groups. Through this traditional way of analyzing data, researchers hope that for the most part they can reject chance as a factor in producing the

difference and instead attribute the difference to the effectiveness of one of the treatments, e.g., a particular program.

Another approach for analyzing data is the concept of **effect size**. This is a measure of practical significance to determine the magnitude of the difference between groups that many evaluators are using (Worthen et al. 1997). Thus effect size indicates the extent to which the difference in an outcome (e.g., test scores) is the result of the intervention given to the experimental group as compared to the method usually applied (i.e., the one that the control group experienced). When evaluating a program, the question raised with respect to effect size is: Has the intervention under study produced a sufficiently large impact so that the difference between groups can be considered worthwhile?

Effect size is symbolized by Δ (for delta). The formula for calculating this statistic is shown next.

$$\Delta = \frac{M_E - M_C}{SD_C}$$

M_E = the mean of the experimental group
M_C = the mean of the control group
SD_C = the standard deviation of the control group[2]

The effect size indicates how much of a **percentile shift** the experimental group has made relative to the control group.[3] In interpreting effect size, the frame of reference for both groups is the 50th percentile point, or the middle of the normal curve. Thus the extent of movement for the experimental group (i.e., how much better it has performed) is measured from this point. For example, suppose that the test performance for students in the innovative and regular program were as follows:

$$\Delta = \frac{70 - 62}{10} = \frac{8}{10} = .80$$

DETERMINING PERCENTILE SHIFT FROM EFFECT SIZE

Range of Effect Size		Corresponding Percentile Shift
From	*To*	
-1.81	-1.70	4
- .82	- .79	21
- .59	- .49	31
- .01	.00	50
.49	.50	69
.79	.82	79
1.70	1.81	96

Figure 8.1

The effect size is then located in Figure 8.1 to determine its corresponding percentile shift. (Not all of the values are shown. See Appendix B for a more complete list.) The shaded row shows the range in which .80 is located. The effect size translates to a percentile shift of 79. It is considered to be a large effect size and represents very good performance. The reason for this is that the innovative group's average test score, which was 70, is now located at the 79th percentile, as compared to the 50th percentile for students in the control group whose average score was 62.

By convention, the criteria shown in Figure 8.2 have been established for interpreting various effect sizes. Note that an effect size of

CRITERIA FOR INTERPRETING EFFECT SIZES

.20 = small

.50 = medium

.80 = large

These three categories were developed by Cohen (1969).

Figure 8.2

1.50 or above can be considered "very large." In instances where the effect size is negative, this means that the control group has made greater gain than the experimental group.

Examining Data Further Using Effect Size Analyses

By disaggregating data, an analysis can reveal if program participants having a certain characterisitc did particularly well (Weiss 1998). One way to determine this is to calculate effect sizes for subgroups of students as shown in Figure 8.3.

Using Percentages to Measure Differences

Depending on the design of the evaluation study, the following types of comparisons can be made to measure differences:

- two-group comparison
- one-group comparison
- one- or two-group comparisons

Each is discussed next.

Two-Group Comparison

To determine if a marked difference has occurred between students participating in the program being evaluated and nonparticipants, first examine the pretest means scores as a check on the equivalency of the groups at a particular point in time (e.g., at the beginning of the school year). These means should be nearly the same. If this is not the case, then the groups should be balanced. This can be done by deleting one or more scores from the top end of the distribution for the group with the higher mean and one or more scores from the low end for the group with the lower mean. Note that the students who are dropped from the pretest should not be counted for the posttest analysis.

DETERMINING THE EFFECTIVENESS OF THE EXPERIMENTAL TREATMENT BY COMPARING SUBGROUPS

Steps:

1. Combine students in the experimental and regular programs and calculate an overall mean based on their pretest scores.
2. Using the overall mean, form two groups by categorizing students according to whether or not they were above average (in the upper half) or below average (in the lower half).
3. Separate students in the upper half of the combined group according to their participation in either the **experimental** or the **regular** program; do the same for students in the lower half of the combined group. There are now four subgroups:

Upper experimental students	Upper regular students
Lower experimental students	Lower regular students

4. Using **posttest scores**, calculate effect sizes for comparisons of the above average experimental students vs. the above average regular students; repeat for the below average students in both groups. Also indicate the corresponding percentile shifts for these effect sizes.
5. Organize the results as shown in the following example.

Group	Effect size	Percentile shift
Above average	.52	70
Below average	.90	82

As shown above, the experimental program had its greatest impact on the below average students since their percentile shift was higher than that for the above average students. We also know that the experimental program was more effective than the regular one because both effect sizes were positive.

Note that a subgroup analysis would be in addition to an overall analysis. Therefore, the results of the comparisons should be displayed in one table as illustrated in this model:

Group	Effect size	Percentile shift
Above average	.52	70
Below average	.90	82
Combined groups	.80	79

A further point: While the example involved posttest achievement scores, an effect size can also be calculated for any quantitative variable of interest (e.g., attitudes toward a subject). Forming subgroups could be done on the basis of a pretest on such attitudes or even by using pretest achievement scores if attitudes as an outcome variable were being compared. Thus the effect size may show that the larger percentile shift occurred in attitudes for those students in the experimental group who were initially low in achievement.

Figure 8.3

In measuring the extent of differences from pre- to posttests, the following approaches can be used:

- If a standardized (or other) test has been administered, the mean gain scores for participants and nonparticipants can be compared. Also report what this percentage represents in terms of improvement. The example shown in Figure 8.4 provides a format for presenting these results.

  To determine the percentage gain, subtract the mean of the pretest from that of the posttest, and divide this difference by the pretest. Thus, in the case shown in Figure 8.4, each student in the innovative program improved an average of 43 percent, while for those in the control group, the gain was only 20 percent.

A COMPARISON OF READING ACHIEVEMENT: STUDENTS EXPOSED TO THE XYZ INDIVIDUALIZED READING PROGRAM VERSUS A BASAL READER APPROACH

Group	N	Pretest Mean	Posttest Mean	Difference Posttest-Pretest	Percent Improvement
XYZ	22	28.20	40.45	12.25	.43
Basal	24	28.40	34.10	5.70	.20

N = number of cases

Figure 8.4

- For a criterion-referenced test, a table can be used to profile the percentage of students who achieve at different levels of mastery on an instrument administered during the evaluation study.[4] Figure 8.5 illustrates a format that can display this information.

  A profile of student achievement as shown in Figures 8.5 is not used to measure gain since no pretest was given. However, a check should be made to determine if the groups were

A PROFILE OF READING ACHIEVEMENT:
STUDENTS PARTICIPATING IN THE XYZ INDIVIDUALIZED
READING PROGRAM VERSUS A BASAL READER APPROACH

Group	N	Did not meet minimum 75% mastery level	Mastered 76-85% of the material	Mastered over 85% of the material
XYZ		N = % =	N = % =	N = % =
Basal		N = % =	N = % =	N = % =

Figure 8.5

comparable in ability at a previous time, e.g., at the beginning of the school year. Grade equivalent scores can be used for this purpose.

One-Group Comparison

In this scenario, it must be determined if a marked improvement has occurred for the group exposed to the program. This approach would

A COMPARISON OF PRE- AND POST-OBSERVATIONS OF DISRUPTIVE
BEHAVIOR FOR A CLASS RECEIVING POSITIVE REINFORCEMENT (N = X)

	Pre observation	Post observation	Difference	Percent Improvement
Mean	xx	xx	xx	xx

Figure 8.6

be used if there was not another group to be compared. For this analysis, the same pre- and postinstruments are needed. Based on the one-group time series experimental design discussed in chapter 4, Figure 8.6 can be used to summarize these results.

One- or Two-Group Comparisons

An approach for a one- or two-group analysis is to compare report card data for two marking periods. These constitute the pre- and post-observation periods. The percentage of students earning A's and B's could, for example, be compared (see Figure 8.7). Note, however, that

A COMPARISON OF REPORT CARD GRADES OVER TWO MARKING PERIODS BETWEEN THE INTERDISCIPLINARY AND REGULAR SOCIAL STUDIES PROGRAMS

Group	Total Number in Program	Percent earning As and Bs		Difference in %
		1st marking period	4th marking period	
Interdisciplinary				
Regular				

Figure 8.7

the students should be the same for the pre- and post-measures. If participants have joined the group after the "pre" period, they should not be included in the analysis. Use the format shown in Figure 8.7 to summarize these analyses. If only one group is participating in the evaluation, the second group would, of course, simply be omitted.

Note that at the time of pretesting, some students may be relatively high in achievement. Therefore, when compared to their posttest

PROCEDURE FOR ANALYZING OBSERVATIONAL DATA

Step:

1. Total the ratings of all items on the instrument assigned by each rater for the first observation and add these values.
2. Divide by 2 to obtain an average for Observation 1.
3. Divide this average by the number of items on the instrument to obtain the mean for the first observation from both raters. The result yields the average overall rating for each item in terms of the five point scale.
4. Repeat steps 1, 2, and 3 for Observation 2. (Repeat again should a third observation be made.)
5. Find the overall mean rating given by both observers for each item by dividing the results from steps 3 and 4 by the number of observations made.

An illustration of this procedure is shown next.

	Observation 1	Observation 2
Total rating from Rater A	56	59
Total rating from Rater B	+ 60	+ 62
Total of ratings from the observation	116	121
Average rating from the observation (÷2)	58	60.5
Average overall rating for each item (÷14)	4.14	4.32

To obtain the overall mean of the items from both observations:

$$\frac{\text{Average item rating from Observation 1} + \text{Average item rating from Observation 2}}{2} \quad \text{or} \quad \frac{4.14 + 4.32}{2} = \frac{8.46}{2} = 4.23$$

On a 5-point scale, this result can be interpreted as a high rating.
For situations in which more than one teacher was observed, the result obtained in step 5 for each teacher would be combined and averaged. For example:

Overall mean rating for each item—Teacher 1	4.10
Overall mean rating for each item—Teacher 2	+4.20
	8.30
	÷ 2
Overall mean ratings for each item—Teachers 1 and 2	4.15

Figure 8.8

performance, they may not have demonstrated as much of a percentage gain as students whose pretest scores were lower. Thus in addition to analyzing gains, another approach is to compare the percentage of students who were above a certain standard (e.g., grade level) on the pre- and posttests.

Analyzing Observational Data

The steps shown in Figure 8.8 can be used to analyze and interpret observational data. The procedure is based on the fourteen item instrument introduced in chapter 7 (Figure 7.4) that uses a five-option continuum to assess student performance. These steps apply to a situation in which classes in an innovative program are being observed. However, the process is repeated if a comparison group is involved.

Prior to making any observations, it would be useful for the raters to review each item on the observation scale and discuss the type of student responses that would constitute the different ratings on the five-option continuum. After the first session, the observers should compare ratings to determine the extent to which they are discrepant. If there are marked differences between them, these should be discussed with a view toward understanding why each made their particular judgment. Comments made in the space provided under each item should be useful for this purpose. Better understanding may reduce differences that could occur on the second observation. If noticeable differences are consistent, this finding should be indicated in the evaluation report since it would lessen the reliability of the observations.

Results of observations can be used for formative and summative purposes. Regarding the latter, the overall score obtained on the rating scale can serve as an indicator of program effectiveness. Formatively, item-by-item analyses could suggest areas for faculty efforts to improve student performance concerning interactions with their teachers and other students. These analyses can contribute

ITEM ANALYSIS FOR OBSERVATIONAL DATA

Steps for analyzing individual items:

1. Rater A: Add the ratings given to an item for each observation made.
   Divide the total by the number of observations.
   This yields the average rating for the item.
2. Rater B: Repeat the process using this observer's ratings.
3. Add the two averages for the item and divide by 2 (the number of raters). This yields the average rating for the item from both raters.

To illustrate:

Item 1	Rater A	Rater B
Observation 1	3	2
Observation 2	+ 4	+ 3
Total	7	5
Average for item from the Rater (Total ÷ 2 observations)	3.5	2.5
Average overall rating for item (Rater A average + Rater B average ÷ 2)	3.5 + 2.5 = 6 ÷ 2 = 3	

Figure 8.9

information concerning the team's recommendations for improving program processes—the diagnostic function of evaluation. An item analysis could also support a favorable assessment by the team as to the program's particular strengths—the confirming function of evaluation. Figure 8.9 illustrates this type of analysis.

Analyzing Portfolios

One method for assessing student work contained in portfolios is to have two members of the evaluation team serve as raters. As found in Gredler (1996), criteria such as the following can be used

collectively to assess student work: self-reflection/self-assessment abilities, thinking processes and understanding (Wolf 1987/1988), as well as creativity (Gearhart, Herman, Baker and Wittaker 1992).

Since portfolios reflect student development (Wolf 1989), if they were kept over the entire school year, improvement should be considered when appraising the work. Whatever the time frame involved, the portfolios should also be judged for overall quality.

Prior to making judgments about each portfolio, the raters should discuss how they would interpret the criteria on the basis of a five-point scale. A separate class mean for improvement and/or one for overall quality should be calculated and considered for contributing evidence to determine program effectiveness. If more than one class in the innovative or regular program is being evaluated, combine the classes for each of the separate groups before calculating their means.

Analyzing Attitudinal Data

As was the case with observational data, the responses to an instrument measuring attitudes toward a subject can be used summatively and formatively. For the latter, an item-by-item analysis can disclose the percentages of students that chose various options on a Likert scale as shown in the following example for fifty students:

The work I do in science is interesting.

	Strongly agree	Agree	Not sure	Disagree	Strongly disagree
N =	3	7	8	22	10
% =	6	14	16	44	20

Disaggregating the data in this way could contribute information for improving the program's operation.

Note that when interpreting an overall **mean** on an attitude questionnaire with a five-point Likert scale format, a judgment should be made in verbal terms that translates what the average response represents. For example, a mean of 4.2 can be categorized as "good" attitudes, but 3.3 as only "fair."

Student opinions about a certain subject reflect attitudes toward it in general. Using a Likert scale, statements referring to the curriculum being evaluated could be added to a student survey and an item analysis applied. The following is an example where the frame of reference is a particular program:

> The honors program in English has helped me improve my writing.

Analyzing Qualitative Data

Qualitative information provides the in-depth dimension of the evaluation study. In analyzing open-ended questions from survey questionnaires or those posed during interviews, search for patterns in the data to determine if there is a majority position. For example, state that eight of the ten parents interviewed felt that the program helped their children develop more interest in reading at home. Next, provide a few representative direct quotes that support the respondents' opinions, even those that may be dissenting.

If a scale that accompanied an interview question contained, for example, the options: very effective, effective, somewhat effective, and not very effective, then the analysis would simply involve counting what the majority of responses were. The next step is to report representative comments.

For focus group interviews, since the opinions of *individuals* are not being tallied with respect to a particular survey question (e.g., how many answered positively), it will be necessary for the interviewers to record their overall impression of the group's responses to the question each interviewer posed. For this purpose, indicate, for example, if all or most of the group answered affirmatively. Also summarize the most typical comments made that represent the group's position.

As stated previously, where applicable, an evaluation question should be answered by obtaining both quantitative and qualitative information. The results of analyses involving both of these sources of data enables a more comprehensive evaluation than if only one approach were used.

PHASE 5—EVALUATING THE PROGRAM'S EFFECTIVENESS AND OFFERING RECOMMENDATIONS FOR ITS FUTURE DEVELOPMENT

In this phase, the team examines all of the evidence and then, through consensus, makes an assessment concerning the program's overall effectiveness, e.g., "very effective" or "somewhat effective." As indicated earlier, the criteria employed in drawing this conclusion are based on professional judgment. These are not stated at the beginning of the evaluation study, but emerge when comparisons are made concerning student progress for program participants and nonparticipants or between pre- and postcomparisons involving a single group. For example, if students in a remedial program make an eight-month gain on a standardized test over those who did not have this experience, how should such improvement be designated? Would the team regard this growth as "marked" or "excellent"? Whatever the label assigned to student development, it should characterize the extent to which the program is meeting its goals.

The exception to waiting for evaluation criteria to emerge is the case when an objective has been formulated at the beginning of the program, as shown in the following example:

> At least 75% of the students participating in the after school reading program will increase their grade equivalent scores by a minimum of six months.

The extent to which an objective such as the preceding has been achieved can be readily measured. As such, it provides more objective evidence to support the team's evaluation.

Considering the Program's Context

A variety of factors found in the setting (i.e., context) in which the program is operating can affect its outcomes in addition to the instructional processes that are part of the curriculum. According to

the Joint Committee on Standards (1994), contextual factors should be acknowledged when interpreting the findings of an evaluation since they can influence the extent to which a program is successful.

A number of these factors were included in the program needs assessment presented in chapter 7. Where responses from items reflect a consensus for "medium" and "high" priority for change, this information can provide a basis for qualifying the overall evaluation by the team of the program's effectiveness. Thus when judgment is less than excellent, it would be tempered by the constraints on the program's operation (e.g., students lacking basic skills in the subject) identified in the needs assessment.

Another issue in evaluating a program is the school's resources—time and financial—that were expended in preparing for and implementing instructional activities.

If an evaluation is qualified in some way, this should be indicated in the report along with a statement of the possible potential of the curriculum for producing better student outcomes when certain program needs are met. Including this position held by the team should enhance the credibility of the report because it is likely to be perceived as a fair appraisal. A further note: The source for qualifying any judgment should not be limited to the information obtained by the program needs assessment. Any questions that tap opinions related to the program processes can be used to identify factors concerning the context in which the curriculum is being delivered.

Considering the Program's Efficiency

Another issue in evaluating a program is the school's resources—time and financial—that were expended in preparing for and implementing instructional activities. This is the efficiency dimension of the evaluation for which the issue of program worth is addressed based on opinions elicited by the following evaluation question (which was raised in chapter 7):

Are the outcomes produced by the program worth the time and/or money expended to prepare for and implement it?

By determining a trend in responses to this question, the team can draw a conclusion about perceptions of the program's worth held by program staff and the administrator(s). The team is then in a position to recommend whether efforts should be directed at finding a program that might be more efficient, that is, one that would not require as many resources to implement. The resource issue is treated in the conclusion section of the evaluation report. Conclusions drawn about a program's effectiveness and efficiency, including recommendations for its improvement, are part of this report and are covered further in chapter 9.

PHASE 6—WRITING THE EVALUATION REPORT

Present the document to stakeholders in two stages: draft and final forms. Consistent with Worthen et al. (1997), for the **draft version**, elicit feedback from stakeholders concerning the quality of the report and any of its aspects that might need to be clarified. Ask respondents for their critical comments about the soundness and usefulness of the report. Consider this feedback carefully for the final version.

One important step is to include an acknowledgment page thanking stakeholders for their input and cooperation in copies of the report that are disseminated to them. As stated earlier, confidentiality and anonymity should be maintained.

Make necessary revisions for the **final report**. Proofread the document and submit it to appropriate audiences. Chapter 9 contains a format for an evaluation report that connects all six phases of the evaluation process.

CHAPTER EIGHT HIGHLIGHTS

1. Quantitative and qualitative data analyses complement each other when conclusions are drawn from the results disclosed by responding to the evaluation questions.
2. An effect size and/or comparisons of percentages can be used to analyze the extent of improvement students have made in analyzing programs, depending on whether one or two groups were involved.
3. Other analyses can be applied to observational and attitudinal data as well as portfolios.
4. Qualitative data provide the in-depth dimension of an evaluation study by obtaining explanations from respondents as to the effectiveness of the innovative program.
5. Through consensus, the evaluation team judges the overall effectiveness of the program and also provides its own and others' recommendations for improvement. This information is part of the evaluation report.

ENDNOTES

1. Hanes and Hail (1999) use meaningful value to the client as one of the indicators for evaluating program outcomes. Along these lines, in discussing criteria for evaluating problems, Worthen et al. (1997) maintain that a standard should reflect a difference in performance that is sufficiently meaningful.

2. The standard deviation, which measures variability of scores from the mean, can be easily determined by using a scientific calculator.

3. The approach used in presenting effect size is based on that found in Suter (1998).

4. For a criterion-referenced test, the proportion of students who meet a standard is one measure that can be used to evaluate programs (Worthen et al. 1997).

BEST COPY AVAILABLE

CHAPTER NINE

WRITING THE EVALUATION REPORT

OVERVIEW

How relevant, credible, and ultimately usable the evaluation is to stakeholders depends in large measure on the quality of the evaluation report.[1] In this regard, based on the works of Owens (1979) and Worthen et al. (1997), this chapter provides a framework for organizing the report to facilitate the writing of a professional document.

The parts of the report to address are shown in Figure 9.1.

CONTENTS OF THE EVALUATION REPORT

I. Title Page
II. Acknowledgments
III. Abstract
IV. Purpose of Evaluation
V. Program Description
VI. Evaluation Questions
VII. Procedures for Collecting and Analyzing Evaluation Information
VIII. Evaluation Results and Interpretation
IX. Conclusions
X. Appendices

Figure 9.1

BEST COPY AVAILABLE

FORMAT AND CONTENT

I. Title Page

An example of a title page is shown in Figure 9.2.

MODEL OF TITLE PAGE

An Evaluation of the _____ Program
at the
_____ School

A Report Submitted by the (year) Evaluation Team

Team Leader: _____
Members: _____
_____
_____
_____

Month, Year

Figure 9.2

II. Acknowledgments

The evaluation team thanks various individuals and groups for their input, cooperation, and/or support of the project.

III. Abstract

The abstract represents a summary (one to two pages) of the entire evaluation report written in the past tense. It is not divided into sections with titles. The following should be included:

A. Indicate that a systematic evaluation was conducted to determine the extent to which the goals of the program were (are being) achieved. (The major goals are stated here.) In the same paragraph, provide a brief description of the program. Also give the time frame during which the study was conducted (e.g., from October through April of the current school year).

B. State that evaluation questions were raised and addressed to guide the project and obtain different types of evidence to determine the program's effectiveness.

C. In three separate paragraphs, summarize sections VII through IX. Tables are not presented.

IV. Purpose of the Evaluation

For an ongoing curriculum, the purpose is two-fold: to determine the effectiveness of the program and to provide recommendations for its improvement. (This perspective is consistent with the benefits described earlier in this book of assessing the outcomes of a program as well as the processes used to deliver its content.)

V. Program Description

Include the description developed in Phase 1—Describing the Program. Specify how it was obtained.

VI. Evaluation Questions

State why these questions were posed. As indicated in chapter 7, they facilitate the team's decisions regarding the types of data to collect, methods used for analyzing the data, and conclusions drawn about the program's impact and methodology. Also, describe the procedures for formulating and selecting evaluation questions. Provide a final list of the questions.

VII. Procedures for Collecting and Analyzing Evaluation Information

A. *Methods of collecting data.* Indicate the time period, that is, the month and year during which data collection occurred. State the number of participants that were involved for each type of data collected. (Include the way sampling was done, if applicable.) Also specify how the data were collected by summarizing the techniques used, e.g., interviews, survey questionnaires and observations. In addition:

1. Provide the sources for test scores (indicate name and/or type of test, a short description, and when taken).
2. Provide the sources for any other type of data obtained from student records.
3. Refer the reader to an appendix to examine a specimen of a particular instrument (other than a standardized test).

B. *Data analysis methods.* Separately describe quantitative and qualitative approaches used to summarize and interpret information. For the quantitative analyses, indicate aspects such as how an attitudinal instrument was tallied and how the performance of the group(s) was determined. For qualitative data, state what trends in responses were identified.

Report the percentage of usable questionnaires that were returned (or collected) based on the number of forms distributed. This percentage provides an indication of how representative the responses are with respect to the population surveyed. If the return rate is relatively low, then any conclusions drawn about opinions should be considered tentative.

VIII. Evaluation Results and Interpretation

A. The section is introduced by informing the reader that the results will be organized according to each evaluation question.[2] Then, state each question followed by a transitional sentence, such as "Information regarding this question is indicated in the following table." Also include the results of the qualitative analyses under the appropriate question.

B. If they occurred, include a discussion of any other positive outcomes or negative aspects of the program. This is the last paragraph in the section.

IX. Conclusions

A. Evaluate the overall effectiveness of the program. For example, state that the program is (was) very effective in achieving its goals. When making any judgment, also describe the criteria that were applied.

B. Discuss the strengths of the program. Indicate which aspects of it should continue as they are currently being implemented. Also state recommendations for improvement. These should include suggestions obtained from stakeholders as well as those offered by the evaluation team. Any that were given by the former should be credited to them. Recommendations should be introduced by a phrase such as

> Three teachers recommended that . . .
> Most parents recommended that . . .
> The evaluation team recommends the following: . . .

(A list then follows.)

Each suggestion offered by the team should be briefly supported by giving the corresponding reason for making it. If the program has been completed, indicate what steps should be taken to improve it (or a similar one) if it were to start up in the future. The evaluation team should also recommend if a program should be terminated and the rationale for this decision.

X. Appendices

Include specimen copies of instruments, e.g., survey questionnaire, interview questionnaire, attitude measure, observation scale, and an achievement test. For each instrument used, include a separate non-standardized appendix with an appropriate title sheet.

WHAT THE EVALUATION REPORT REPRESENTS

The report is the culminating product of an evaluation study. As such, it should not only be clearly written (Joint Committee on Standards 1994; Weiss 1998; Worthen et al. 1997), but it should also provide a basis to support the conclusions of the evaluation team. This will contribute to the report's acceptance and usefulness. To achieve these ends, the report should highlight the two basic functions of program evaluation stated in chapter 1: *confirming* the extent to which a program is effective and *diagnosing* areas in which improvements are needed. The reasons why these functions are pivotal are discussed next.

- **Confirming.** When a program evaluation identifies outcomes indicating how those implementing the curriculum have been successful, this serves as recognition from peers—an important source of gratification for professionals.
- **Diagnosing.** Based on objectively analyzed evidence,[3] promising modifications that are offered to improve areas of the curriculum provide direction for future delivery of the program. Such suggestions are likely to be perceived as useful since they could facilitate the teaching/learning process. Moreover, the report should be more readily accepted since it would also contain recommendations given by those implementing the program. Seeking their input is in keeping with the need for stakeholders to feel ownership in the evaluation study as advocated by Worthen et al. (1997).

In essence, the report should be respected for its excellence, fairness,[4] and professionalism by those who are responsible for implementing the curriculum within a school, by administrators in the district, and by other stakeholder groups—parents and community members—whose interest in and support of school efforts are vital.

As indicated earlier, program evaluation conducted by practitioners in a local school involves reflective practice. This applies to the evaluators, as well as those who carry out and administer the program as each group engages in thought and dialogue about instructional activities and implications of student outcomes. As a result of this process, the evaluation report should represent a viable reference from which administrators and faculty (including future evaluation team members) can learn. This perspective is consistent with the tenets of a learning organization.

In this context, writing the report can also serve as a professional development activity for the evaluation team as it proceeds in addressing the contents of the document. The checklist shown in Figure 9.3 can be used as a guide in completing the report.[5]

EVALUATION REPORT CHECKLIST

PART I: REPORT CONTENTS

I. Title Page

- ❑ Does the title page contain the type of information shown in the model (Figure 9.2)?

II. Acknowledgments

- ❑ Have all individuals and groups involved been mentioned?

III. Abstract

A. Have the following been indicated?
- ❑ That a systematic evaluation was conducted
- ❑ The extent to which goals were (or are being) achieved
- ❑ The major goals
- ❑ A brief description of the program
- ❑ The time frame for the study

B. Were evaluation questions raised for the following?
- ❑ To guide the project
- ❑ To obtain different types of evidence to determine program effectiveness

C. Three separate paragraphs should summarize the following (which correspond to sections VII, VIII, and IX of the report, respectively).
- ❑ Procedures for collecting and analyzing evaluation information
- ❑ Evaluation results and interpretation
- ❑ Conclusions

IV. Purpose of the Evaluation

Have the following purposes of the evaluation been stated?
- ❑ To determine the effectiveness of the program
- ❑ To provide recommendations for its improvement

V. Program Description

- ❑ Has the program description been included?
- ❑ Was information given on how the description was obtained?

VI. Evaluation Questions

- ❑ Were the procedures for formulating and selecting evaluation questions described?
- ❑ Was a final list of questions included?

Figure 9.3

VII. Procedures for Collecting and Analyzing Evaluation Information

A. Methods of collecting data
- ❑ Was the time period during which data collection occurred indicated?
- ❑ Was the number of participants given for each type of data collected?
- ❑ Was the sampling method described (if applicable)?
- ❑ Were data collection techniques summarized?
- ❑ Were the sources for test scores and/or any other type of data obtained from records provided?
- ❑ Has the reader been referred to an appendix for a specimen copy of the instrument?

B. Data analysis methods
- ❑ Were there separate descriptions for the quantitative and qualitative approaches used to summarize and interpret information?

VIII. Evaluation Results and Interpretation
- ❑ Was the reader informed that the results were organized according to each evaluation question?
- ❑ Was each question stated and the results explained?
- ❑ Have technical concepts (e.g., effect size, percentile shift) been explained clearly?
- ❑ Has the return rate for usable survey questions been indicated?
- ❑ If they occurred, were any other positive or negative aspects of the program discussed?

IX. Conclusions
- ❑ Was the overall effectiveness of the program evaluated?
- ❑ Were the criteria used to judge overall effectiveness described?
- ❑ Were the strengths of the program discussed?
- ❑ Were conclusions drawn about the program's efficiency?

A. Have the following been indicated?
- ❑ Aspects of the program that should continue as they are currently being implemented
- ❑ Recommendations for improvement
  - ❑ Suggestions obtained from stakeholders and credited to them
  - ❑ Suggestions offered by the evaluation team including a brief reason for each

B. If applicable:
- ❑ Have steps been indicated to improve a completed (or similar) program if it were to start up in the future?
- ❑ Has the rationale for the team's recommendation for terminating a program been included?

Figure 9.3 (continued)

X. Appendices

- ❑ Have specimen copies of instruments been included, each as a separate appendix with a title sheet?

PART II: MECHANICS

A. Is the information presented accurate? For example, have the following been checked?
   - ❑ Number of participants
   - ❑ Testing dates
   - ❑ Titles of texts used
   - ❑ Scheduling of classes
   - ❑ Results of analyses

B. Has the report been proofread by the evaluation team for the following?
   - ❑ Content
   - ❑ Omissions
   - ❑ Grammatical errors
   - ❑ Typographical errors

C. Overall, does the document have a professional appearance?
   - ❑ Has an attractive font been selected?
   - ❑ Does the use of spacing and headings enhance readability?
   - ❑ Have the formats for the tables been followed?
   - ❑ If graphics have been included, do they contribute to the presentation?
   - ❑ Are the cover and binding appropriate for a professional document?

Figure 9.3 (continued)

CHAPTER NINE HIGHLIGHTS

1. How relevant, credible, and ultimately usable the evaluation is to stakeholders depends in large measure on the quality of the evaluation report.
2. A format for systematically organizing the evaluation report should be followed.
3. The report should confirm the extent to which a program is effective and also diagnose areas in which improvements are needed.
4. The report should be respected for its excellence, fairness, and professionalism.
5. The report should represent a viable reference from which administrators and faculty can learn.
6. Writing the report can serve as a professional development activity.
7. Using a checklist that addresses the contents and mechanics of a report can facilitate the writing of it.

ENDNOTES

1. From their examination of the literature, Worthen et al. (1997) discuss credibility, relevance, and other factors that relate to the use of the evaluation report.

2. This approach is recommended by Worthen et al. (1997).

3. That objectivity should be a characteristic of program assessment is consistent with reporting evaluations impartially (Joint Committee on Standards 1994; Worthen et al. 1997).

4. The importance of fairness in evaluation reports is underscored by Worthen et al. (1997) who regard this standard as the hallmark of professionalism.

5. This checklist, in part, builds upon the works of the Joint Committee on Standards (1994) and Worthen et al. (1997). The latter authors have also created a useful checklist for writing the evaluation report.

BEST COPY AVAILABLE

CHAPTER TEN

ASSESSING THE EVALUATION PROJECT

OVERVIEW

Formative and summative approaches to program evaluation can also be applied to a metaevaluation, which involves assessing the evaluation study itself (Joint Committee on Standards 1994; Worthen et al. 1997). This chapter describes the purpose of each of these approaches and how to obtain (a) formative information to improve the operations of an evaluation currently in progress and (b) summative information to guide future program assessments.

PURPOSES OF METAEVALUATION

The formative role played in carrying out a metaevaluation involves guiding "the planning and implementation of a program evaluation" (Joint Committee on Standards 1994, p. 185). However, "the summative role is to assess the worth and merit of a completed program evaluation" (p. 185).

Conducting both aspects of metaevaluation by the evaluation team would be a productive endeavor since these activities "can be expected to facilitate early identification and corrections of potentially fatal

BEST COPY AVAILABLE

flaws in a program evaluation plan, increase the likelihood that evaluation conclusions are valid, and enhance the acceptance by users and stakeholders, and . . . may increase the cost effectiveness of the program evaluation" (Joint Committee on Standards 1994, p. 186).

FORMATIVE METAEVALUATION

Worthen et al. (1997) indicate the importance of monitoring the evaluation while it is in progress so that revisions can be made in light of any problems encountered. This activity is consistent with the formative aspect of program evaluation. In this regard, the evaluation team should address the question of whether the project tasks are being completed as planned (Worthen et al. 1997). If they are not, asking what improvements or changes are needed to remedy the situation should be elicited. Such an inquiry should be made at each phase of the evaluation project. The form shown in Figure 10.1 can be used to facilitate this task.

Applying a formative approach to identifying and improving aspects concerning the process of program evaluation should also extend to how well the evaluation team is functioning. For this purpose, its members should respond to and discuss questions about the effectiveness and efficiency of the team process. This can be done at a meeting with a view toward exploring actions that can be taken to improve any aspect of how the team is operating. Figure 10.2 lists examples of formative questions that can be raised.

MONITORING THE EVALUATION PROJECT

Phase	Task	Task Completed Successfully Yes	No	If "no," indicate action(s) needed.
1. Describing the program	Write a description of the program.			
2. Providing direction for the evaluation	Formulate evaluation questions.			
	Decide who qualify as stakeholders.			
	Distribute program goals and evaluation questions to stakeholders.			
	Ask stakeholders for suggestions for other questions.			
	Stakeholders and evaluation team judge which questions are addressed.			
	Obtain consensus between evaluation team and stakeholders on final choice of questions.			

Figure 10.1

Phase	Task	Task Completed Successfully Yes	No	If "no," indicate action(s) needed.
Before collecting any data, complete the Evaluation Plan (Figure 7.5 in chapter 7).				
3. Obtaining information to answer evaluation questions	Decide whether to use a sample or entire population of students.			
	Use sampling (if applicable).			
	Decide if a control group should be used.			
	Select data collection methods for each evaluation question: • unobtrusive measures • survey questionnaire(s) • interview questions • follow-up questions • obtaining process information • observations (if conducted)			
	Adhere to ethical data collection methods.			

Figure 10.1 (continued)

SkyLight Professional Development

Phase	Task	Task Completed Successfully Yes	No	If "no," indicate action(s) needed.
4. Analyzing data to assess the program's impact	Appropriate for the evaluation question: • complete analysis of quantitative data • complete analysis of qualitative data			
5. Evaluating the program's effectiveness and offering recommendations for its future development	Draw conclusions about the program's effectiveness and efficiency.			
	Provide recommendations for the program's improvement.			
6. Writing the evaluation report	Present draft version of document to stakeholders for feedback.			
	Make necessary revisions for the final report.			

Figure 10.1 (continued)

Skylight Professional Development

180

JUDGING EVALUATION TEAM FUNCTIONING: QUESTIONS FOR FORMATIVE ASSESSMENT OF EVALUATION TEAM PROCESS

- As a team, are we accomplishing our tasks well?
- Are we on schedule according to the Evaluation Time Log (Figure 7.6 in chapter 7)?
- Are our meetings efficient with little or no wasted time?
- Do you look forward to attending our meetings?
- Do you have any recommendations for improving the team process?

Figure 10.2

The types of questions in Figure 10.2 are compatible with the spirit of a collaborative learning culture. From this perspective, they could facilitate introspection among the group, thereby leading to personal and professional growth.

SUMMATIVE METAEVALUATION

The questionnaire in Figure 10.3 is an example of an instrument that can be used by the evaluation team in its overall assessment of the evaluation project after it has been completed. It should be answered anonymously and the results summarized in a report to the principal by the team leader. This activity should be helpful to the succeeding team. Items can be added, modified, or omitted.

JUDGING THE EVALUATION PROJECT

Evaluation Project Assessment Questionnaire

The following items reflect the work of the assessment team as well as your role in the program evaluation project. Your responses will provide useful information for the next evaluation team.

Directions

Please rate each item according to the scale below it by circling **one** number that best indicates your judgment. Any comments to explain your opinion will be very helpful.

1. Plans for conducting the evaluation were sound.*

   1 2 3 4 5
   Low High

   Comment:_____

2. Findings of the evaluation were clear.**

   1 2 3 4 5
   Low High

   Comment:_____

3. Conclusions were justified by the data.**

   1 2 3 4 5
   Low High

   Comment:_____

4. Procedures were carried out as planned.*

   1 2 3 4 5
   Low High

   Comment:_____

5. The evaluation report reflects professional quality.

   1 2 3 4 5
   Low High

   Comment:_____

6. The report was disseminated according to schedule.*

   1 2 3 4 5
   Low High

   Comment:_____

*Items 1, 4, and 6 are based on Worthen et al. (1997).
**Items 2 and 3 are based on the Joint Committee on Standards (1994).

Figure 10.3

7. The report was well received by stakeholders.

   1 2 3 4 5
   Low High

   Comment:_____

8. Feasible follow-up plans were made with the principal and program staff to modify the curriculum.

   1 2 3 4 5
   Low High

   Comment:_____

9. My learning as a professional has increased by having been a member of the evaluation team.

   1 2 3 4 5
   Low High

   Comment:_____

10. Participating in the evaluation project was worth the time and effort since the assessment and recommendations can help improve student achievement.

    1 2 3 4 5
    Low High

    Comment:_____

11. Do you have any recommendations for improving the process of conducting an evaluation for the next evaluation team?

    Comment: _____

    _____

    _____

    Thank you for your input.

Figure 10.3 (continued)

A CONCLUDING NOTE

Metaevaluation and reflective practice are congruous. Thus carrying out each phase of metaevaluation requires reflection. This occurs at the planning, process, and terminal stages of the study. As with any endeavor to improve a situation, current efforts should be analyzed, alternatives weighed, a plan implemented, and its effects assessed. In this regard, metaevaluation strengthens the link between the process of assessing a program and its development since sound evaluation discloses where and how future instruction should be concentrated.

CHAPTER TEN HIGHLIGHTS

1. Formative and summative approaches can be applied to assessing the evaluation study itself. These activities constitute a metaevaluation.
2. Formative information should be obtained at each phase of the evaluation project. This task can be facilitated by using a checklist to make any needed modifications.
3. Questions can be raised to determine the effectiveness and efficiency of the team process.
4. A questionnaire answered by the team members to measure their overall assessment of the completed evaluation project should be helpful to succeeding teams.
5. Metaevaluation and reflective practice are congruous since reflection is required when carrying out this type of evaluation.
6. Metaevaluation strengthens the link between program assessment and its development.

BEST COPY AVAILABLE

CHAPTER ELEVEN

REVISITING THE PRINCIPAL'S LEADERSHIP ROLE IN PROGRAM EVALUATION

OVERVIEW

This chapter expands on the basic premise of the book: conducting a program evaluation that leads to higher student achievement depends, in the final analysis, on what the principal has done to promote and support the various phases of the project. An instrument for a principal's self-assessment is also included to raise into saliency leadership factors to help ensure that the evaluation process makes a meaningful impact on the organization. The chapter concludes with a perspective on winning that views standardized testing as a welcome opportunity to demonstrate effective teaching and learning.

THE MULTIDIMENSIONAL RESPONSIBILITIES OF THE PRINCIPAL IN PROGRAM EVALUATION

This book has described a systematic and comprehensive approach for conducting a program evaluation. However, the potential of an assessment study for improving student achievement will not be realized without the leadership initiatives[1] and follow-through actions a principal can take to help ensure the project's success. In this regard,

BEST COPY AVAILABLE

DeRoche (1987) argues that "a principal's attitudes and procedures can make the evaluation program an enlightening, interesting, exciting venture or one that is frustrating, based on fear, and viewed negatively by those involved" (p. 6).

Considering this perspective, specific implications of a principal's active intervention in various phases of an assessment project are indicated next.

- When a principal promotes the need for an evaluation team, supports its activities, and provides a place for it as a regular part of the school organization, this speaks to the importance of assessing programs as a means of improving student achievement.
- When a principal makes her- or himself available to assist the team by handling administrative type tasks if called upon, this facilitates the completion of the project.
- When a principal acts on the recommendation of the evaluation team at the close of the study, this helps ensure that its report will be used to enhance program delivery.
- When a principal recognizes the team's accomplishments as well as those of the program staff and their students, this contributes to feelings of satisfaction and heightened motivation.

The principal's involvement in an evaluation project is driven by the attitudes he or she has toward this endeavor. Encouraging attitudes find expression when the administrator communicates the benefits that can occur as a result of the faculty conducting its own program assessments.

ATTITUDES OF PRINCIPALS TOWARD ASSESSING PROGRAMS

Standardized tests provide one gauge for measuring organizational performance. However, as indicated earlier, the results they yield should not be the only means for judging a program's effectiveness

(Joint Committee on Standards 1994; Worthen et al. 1997). Rather, this determination should be based on a number of indicators since these provide additional evidence to corroborate the extent to which a program is successful. Standardized tests obviously would not be used to judge student achievement where there are a variety of curricular goals, e.g., interpersonal relations, community service, and creativity in science and social studies projects. Therefore these types of outcomes should also be assessed to determine the overall worth of a program if they reflect its goals. Put another way, if a goal is important enough to include in a program, the extent to which it has been achieved must be considered when drawing conclusions about the program's impact.

The principal's involvement in an evaluation project is driven by the attitudes he or she has toward this endeavor.

With this in mind, a principal should convey to the staff the value of assessing program outcomes and processes in explicit terms. The following are recommended points to cover.

Evaluating program outcomes in a comprehensive way is worthwhile.

It provides an overall perspective on the extent to which students are progressing on each curriculum goal. As such, when evidence of success is found, this confirms for members of a learning organization where efforts have been effective. If assessment discloses results that are less than satisfactory, then this information can help focus attention on areas needing remediation.

Determining a program's success solely on the results of standardized tests is not evaluation, it is measurement—professional judgment is not involved.

Test scores can have the effect of driving the curriculum because they provide a narrow focus on what is being assessed (Joint Committee on Standards 1994). This effect can detract educators from concentrating on other important goals besides those that the tests measure.

Using valid indicators of program effectiveness helps gauge the level of professional performance.
Since each school is accountable for student achievement, it is essential for the staff, as professionals, to use valid indicators to judge their work systematically and comprehensively. Furthermore, program evaluation should also be included as part of the curriculum development process.

Reflecting on practice suggests areas in need of improvement.
The process of assessing programs contributes to professional development. Since this process requires reflecting on practices, it serves as a stimulus for thought on how to improve curriculum delivery. Thus such reflection involves a level of introspection, both individual and group, that can lead to promising strategies in areas of needed growth for educators.

BEYOND THE TECHNICAL

Program evaluation should be considered a tool that can be applied to enhance organizational outcomes. As such, it yields information that leads to decisions regarding directions for improving curricular activities that may elevate student achievement. This information also suggests areas for the principal to focus on in exercising instructional leadership. (One source for this data is the needs assessment completed by program staff that was presented in chapter 7.)

Undergirding what any curriculum can accomplish are inner qualities of a principal's leadership such as conviction, confidence, passion, resolve (Jason 1998), optimism, and spirit. These are intangible, yet powerful factors that drive endeavors because they galvanize followers.

These factors provide the emotional fuel that propels an organization because they inspire followers as well as leaders themselves to take actions toward higher levels of accomplishment. Such factors work their influence because people function better in an environment

where their energies are not dissipated by negativism, worry, and/or pessimism. Positive emotions that are reflected in a leader's demeanor and behavior communicate the promise of a better quality of life.[2] In this regard, people in an organization become less daunted by difficulties because they see that their leader is at least trying to remedy them and can possibly do so. In fact, "difficulties" cast in a different light and perceived as challenges are motivating. It is in the human psyche to rise to a challenge, and this tropism can be channeled toward eliminating or minimizing barriers that hamper better student achievement.

The Potential of Schoolwide Leadership

While this chapter has focused on the role of the principal in program evaluation, no one person should be thought of as the fountainhead for leadership actions. This is consistent with Schmoker (1999) who indicates that only rare principals can, by themselves, create higher achievement. He points out that the potent combination of principal and teacher leadership increases the chance for instructional changes to occur. Schmoker further maintains that any of the staff can exercise leadership by assuming responsibility for the organization's success by influencing other members.

Against this background, principals should communicate the potential that leadership initiatives taken by all staff members have in improving school performance. When this mentality is pervasive, student achievement should be raised since the collective weight of experience, expertise, and commitment is brought to bear on solving problems—a characteristic of learning organizations.[3] Thus, consistent with DuFour and Eaker (1992), when personnel in a school perceive of themselves as learners as well as professionals, they become more willing to experiment with and modify programs as needed. In this context, these authors refer to Barth's (1990) concept of a "community of learners" in which commitment to experimentation and school improvement are linked. Moreover, in keeping with what Barth

"Difficulties" cast in a different light and perceived as challenges are motivating.

advocates, members of a learning organization embrace the belief that learning can be exciting and rewarding, not only for their students but for themselves as well.

PRINCIPALS' SELF-ASSESSMENT

With respect to the principal's role in program evaluation, the instrument in Figure 11.1 should be useful in guiding her or his actions in the assessment process. Note that most questions can be answered during the course of the evaluation project, and any changes in actions can be made to facilitate the work of the evaluation team. Principals can also reflect on their role after an evaluation has been completed, and, as members of a learning organization, they can apply what they have gained through this self-assessment for future projects.

A CONCLUDING NOTE

At the beginning of this work it was stated that this book was about winning. But this perspective goes beyond the traditional notion that if there are winners, there must also be losers. Instead, consider the potential that winning can have if self-improvement were emphasized. This outlook can lead to students achieving at higher levels than would be the case if the focus were on informing them about how they stood relative to a grade level standard. Athletes know well the value of judging their performance in terms of how they are doing now as compared to what they have demonstrated earlier. When they experience a "personal best," this carries with it feelings of excitement and satisfaction as well as increased motivation to exceed this level. This experience should also occur in an academic setting. Thus, according to Schmoker (1999), teachers should discuss with their students on a regular basis the progress they are making and also

Personal mastery is winning on a personal level.

SELF-RATINGS OF THE PRINCIPAL'S ROLE IN PROGRAM EVALUATION

PRINCIPAL SELF-ASSESSMENT QUESTIONNAIRE: PROGRAM EVALUATION PROJECT

	Yes	Somewhat	No
1. Have I promoted a sufficient need for program evaluation?*	❑	❑	❑
Specific action to take:			
_____			
_____			
2. Have I followed systematic procedures when forming the evaluation team?	❑	❑	❑
Specific action to take:			
_____			
_____			
3. Have I provided resources to facilitate the completion of the evaluation project?*	❑	❑	❑
Specific action to take:			
_____			
_____			
4. Have I shown enthusiasm for the evaluation project?			
Specific action to take:	❑	❑	❑
_____			
_____			
5. Have I been available for assistance, if needed, by the evaluation team?	❑	❑	❑
Specific action to take:			
_____			
_____			

*Items 1 and 3 are based on recommendations for the principal in the evaluation process indicated by DeRoche (1987).

**Although not in the context of conducting program evaluations, Schmoker (1999) discusses the importance of leaders giving public and private recognition for school improvement. This is reflected in items 7, 8, and 9.

Figure 11.1

	Yes	Somewhat	No
6. Have I encouraged the evaluation team?	❑	❑	❑
Specific action to take:			
7. Have I formally and informally recognized accomplishments of the program faculty being evaluated?**	❑	❑	❑
Specific action to take:			
8. Have I formally and informally recognized accomplishments of the students in the program being evaluated?**	❑	❑	❑
Specific action to take:			
9. Have I formally and informally recognized accomplishments of the evaluation team?**	❑	❑	❑
Specific action to take:			
10. Have I followed up on recommendations from the evaluation team?	❑	❑	❑
Specific action to take:			

Figure 11.1 (continued)

recognize improvement and effort. Based on this point, the goal would be to have students develop enough confidence in their abilities so that they feel they can earn satisfactory or even excellent scores on standardized tests. This belief, coupled with having both sufficient academic preparation and test-taking skills, should lessen stress and debilitating anxiety that can result from pressure to do well. Pressure to perform in any arena can in the long run be counterproductive, especially if it exceeds a tolerable point. Thus, ironically, exhortations to achieve are not likely to have their intended effect. Exacerbating this is students and teachers knowing the consequences of substandard results. This environment makes a prospective test event, as well as the testing situation itself, too intensive and detracts from making school a satisfying and enjoyable experience. To counteract this requires a diametrically different view of standardized testing, from that of a threat to an event that provides a good opportunity to demonstrate how successful students and teachers have been in performing their roles.[4] A positive perception of an upcoming challenging situation is liberating and can help minimize the deleterious effects that excess worrying has on preparing for and taking tests.

Considering testing as an opportunity is congruous with conducting program evaluation with a view toward self-improvement, not only for students but for faculty as well. When students meet or exceed standards in various areas of knowledge and skills, this adds information to the outcomes dimension of evaluation. However, where their performance is less than satisfactory, such results will, as stated earlier, suggest directions for improving the process of delivering the curriculum. Assessing this process fits within the framework of personal mastery that for Senge (1990) is pivotal in a learning organization. Personal mastery is winning on a personal level. In terms of school performance, this perspective is practical because it extends to helping students recognize where they need to improve. It enables the program staff to be more efficient in remedying aspects of the curriculum and its delivery because such efforts can then be concentrated on where there is greater need. Therefore, focusing on deficiencies which performers in various disciplines accept as a given, is essential not only for improvement, but for achieving excellence as well.

Consistent with Senge (1990), the concept of continuous self-improvement is inherent in the thrust of a learning organization. When personnel in a school adopt the values of such an organization, they will more readily embrace the promise of lessons that can be learned from program evaluation projects.

In bringing this book to a close, it seems appropriate to point out that individuals achieve more when they accept a message from those they believe in that conveys confidence, support, and optimism about their future performance.[5] Therefore, administrators and staff should communicate the following sentiment: *Bring on the tests. Our students are ready for them, and they will succeed.*

Conducting program evaluations by local school personnel will help strengthen this conviction by guiding actions aimed at preparing students to do well not only on standardized tests but also on other indicators of effectiveness. In this context, the professional knowledge gained by educators from assessing their work can provide the momentum to raise achievement to new heights.

CHAPTER ELEVEN HIGHLIGHTS

1. The potential of a program evaluation for improving student achievement will not be realized without the principal's leadership initiatives and follow-through actions.
2. The principal's involvement in an evaluation project is driven by the attitudes he or she has towards this endeavor.
3. A principal should convey to the staff the value of assessing program outcomes and processes in explicit terms.
4. Undergirding what any curriculum can accomplish are inner qualities of a principal's leadership.
5. Leadership initiatives taken by all staff members can improve the performance of the school as a learning organization.
6. Self-assessment by principals regarding their role in program evaluation should be useful in guiding their actions in the assessment process.
7. If winning were emphasized from a self-improvement perspective, this could lead to higher levels of student achievement.
8. Standardized testing should be viewed not as a threat, but as an opportunity to demonstrate successful teaching and learning.
9. A positive perception toward testing is liberating and can help minimize the effects of worry on preparing for and taking tests.
10. Viewing winning as personal mastery helps students recognize what they need to improve and enables program staff to concentrate their efforts on where there is greater need.
11. Administrators and staff should communicate the sentiment that their students are ready for testing and will be successful.

BEST COPY AVAILABLE

ENDNOTES

1. While his book encompasses both program and personnel evaluation, DeRoche (1987) identifies the principal's formulation of evaluation plans and procedures as a leadership activity.

2. Based on Wheatley's book *Leadership and the New Science*, published in 1994, Schmoker (1999) discusses how destructive thoughts that detract from improving a workplace can be replaced by school leaders who promote a positive, productive, and purposeful culture by celebrating accomplishments.

3. The link between collective intelligence and school improvement is underscored by Schmoker (1999).

4. In discussing evaluation, DeRoche (1987) indicates that the process should help determine the success of teaching and learning.

5. In this vein, Schmoker (1999) calls attention to the need for leaders to communicate confidence regarding what can be achieved by the organization.

APPENDIX A

REPRESENTATIVE SOURCES FOR AUTHENTIC ASSESSMENT

Barton, J., and A. Collins, eds. 1997. *Portfolio assessment: A handbook for educators.* New York: Addison-Wesley.

Belanof, P., and M. Dickson. 1991. *Portfolios: Process and product.* Portsmouth, NH: Heinemann.

Burke, K. 1999. *The mindful school: How to assess authentic learning.* Arlington Heights, IL: SkyLight Training and Publishing.

Burke, K., R. Fogarty, and S. Belgrad. 2002. *The portfolio connection: Student work linked to standards.* Arlington Heights, IL: SkyLight Professional Development.

Darling-Hammond, L., J. Ancess, and B. Falk. 1995. *Authentic assessment in action: Studies of schools and students at work.* New York: Teachers College Press.

Gredler, M. E. 1999. *Classroom assessment and learning.* New York: Longman.

Hart, D. 1994. *Authentic assessment: A handbook for educators.* New York: Addison-Wesley.

Janesick, V. J. 2001. *The assessment debate: A reference handbook.* Santa Barbara, CA: ABC-CLIO.

Kerka, S. 1995. *Techniques for authentic assessments: Practice application brief.* Washington, DC: Office of Educational Research and Improvement.

Montgomery, K. 2001. *Authentic assessment: A guide for elementary teachers.* New York: Addison Wesley Longman.

Seidel, S., and J. Walters. 1997. *Portfolio practices: Thinking through the assessment of children's work.* Washington, DC: National Education Association.

Taggart, G., D. Phifer, J. Nixon, and M. Wood. 1998. *Rubrics: A handbook for construction and use*. Lancaster, PA: Technomic.

Tierney, R., M. Carter, and L. Desai. 1991. *Portfolio assessment in the reading-writing classroom*. Norwood, MA: Christopher-Gordon.

Wiggins, G. 1990. *The case for authentic assessment*. Washington, DC: ERIC Clearinghouse on Tests, Measurement, and Evaluation. (ERIC Document Reproduction Service No. 328 61!)

APPENDIX B

RANGES OF EFFECT SIZE AND CORRESPONDING PERCENTILE SHIFTS

To determine the percentile shift for the experimental group, locate the range in which the calculated effect size falls. (In determining ranges of effect size, values from the normal curve table were rounded to two places. Corresponding percentile shifts are reported in whole numbers.)

Range of Effect Size		Corresponding Percentile Shift
From	*To*	
< -2.32	-2.18	1
-2.17	-1.97	2
-1.96	-1.82	3
-1.81	-1.70	4
-1.69	-1.60	5
-1.59	-1.52	6
-1.51	-1.44	7
-1.43	-1.38	8
-1.37	-1.32	9
-1.31	-1.26	10
-1.25	-1.21	11
-1.20	-1.16	12
-1.15	-1.11	13
-1.10	-1.06	14
-1.05	-1.02	15
-1.01	- .98	16
- .97	- .94	17
- .93	- .90	18
- .89	- .86	19
- .85	- .83	20
- .82	- .79	21
- .78	- .76	22

Range of Effect Size		Corresponding Percentile Shift
From	*To*	
- .75	- .73	23
- .72	- .70	24
- .69	- .66	25
- .65	- .63	26
- .62	- .60	27
- .59	- .57	28
- .56	- .54	29
- .53	- .52	30
- .51	- .49	31
- .48	- .46	32
- .45	- .43	33
- .42	- .40	34
- .39	- .38	35
- .37	- .35	36
- .34	- .32	37
- .31	- .30	38
- .29	- .27	39
- .26	- .25	40
- .24	- .22	41
- .21	- .19	42
- .18	- .17	43
- .16	- .14	44
- .13	- .12	45
- .11	- .09	46
- .08	- .07	47
- .06	- .04	48
- .03	- .02	49
- .01	.01	50
.02	.03	51
.04	.06	52
.07	.08	53
.09	.11	54
.12	.13	55
.14	.16	56
.17	.18	57
.19	.21	58
.22	.24	59
.25	.26	60
.27	.29	61
.30	.31	62
.32	.34	63
.35	.37	64

Range of Effect Size		Corresponding Percentile Shift
From	*To*	
.38	.39	65
.40	.42	66
.43	.45	67
.46	.48	68
.49	.50	69
.51	.53	70
.54	.56	71
.57	.59	72
.60	.62	73
.63	.65	74
.66	.69	75
.70	.72	76
.73	.75	77
.76	.78	78
.79	.82	79
.83	.85	80
.86	.89	81
.90	.93	82
.94	.97	83
.98	1.01	84
1.02	1.05	85
1.06	1.10	86
1.11	1.15	87
1.16	1.20	88
1.21	1.25	89
1.26	1.31	90
1.32	1.37	91
1.38	1.43	92
1.44	1.51	93
1.52	1.59	94
1.60	1.69	95
1.70	1.81	96
1.82	1.95	97
1.96	2.16	98
2.17	2.35 >	99

Note: Any effect size above 2.17 is reported at the 99th percentile since a group scoring this high cannot exceed itself. (See Ravid 1994, on this point in her discussion of percentile ranks.) Conversely, effect sizes of -2.32 are reported as a percentile shift of 1 since in terms of whole numbers, the group would be at the 1% level.

BIBLIOGRAPHY

Barth, R. 1990. *Improving schools from within*. San Francisco: Jossey-Bass.

Beane, J. A., ed. 1995. *Toward a coherent curriculum. 1995 yearbook of the Association for Supervision and Curriculum Development*. Alexandria, VA: Association for Supervision and Curriculum Development.

Brainard, E. A. 1996. *A hands-on guide to school program evaluation*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Burns, J. M. 1978. *Leadership*. New York: Harper and Row.

Campbell, D. T., and J. C. Stanley. 1963. *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Combs, A. W., and D. Snygg. 1959. *Individual behavior*. New York: Harper and Row.

Cohen, J. 1969. *Statistical power analyses for the behavioral sciences*. New York: Academic Press.

Csikszentmihalyi, M. 1990. *Flow: The psychology of optimal experience*. New York: Harper Perennial.

DeRoche, E. F. 1987. *An administrator's guide for evaluating programs and personnel: An effective schools approach* (2nd ed.). Boston: Allyn and Bacon.

DuFour, R., and R. Eaker. 1992. *Creating the new American school: A principal's guide to school improvement*. Bloomington, IN: National Education Service.

Gearhart, M., J. Herman, E. Baker, and A. Wittaker. 1992. *Writing portfolios: Potential for large-scale assessment*. Los Angeles: Center for the Study of Evaluation, University of California.

Gredler, M. E. 1996. *Program evaluation.* Englewood Cliffs, NJ: Prentice-Hall.

Hanes, J. C., and M. Hail. 1999, November. *New indicators for program evaluation.* Paper presented at the Annual Meeting of the American Evaluation Association, Orlando, FL. (ERIC Document Reproduction Service No. ED 437 412)

Hasenstab, J. K., and C. C. Wilson. 1989. *Training the teacher as a champion.* Nevada City, CA: Performance Learning Systems.

Haertel, E. 1992. Performance measurement. In *Encyclopedia of educational research,* edited by M.C. Alkin, pp. 984–989. New York: Macmillan.

Hopkins, K. D., and G. V. Glass. 1978. *Basic statistics for the behavioral sciences.* Englewood Cliffs, NJ: Prentice-Hall.

Jason, M. 1998. Perspectives on the principal's resolve as a factor in creating the quality school. In *Creating high functioning schools: Practice and research,* edited by Y. Cano, F. H. Wood, and J. C. Simmons, pp. 21–28. Springfield, IL: Charles C. Thomas Publisher.

Jenkins, J. A. 1993, April. *Can quality program evaluation really take place in schools?* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA. (ERIC Document Reproduction Service No. ED 397 067)

Joint Committee on Standards for Educational Evaluation. 1994. *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Kanter, R. M. 1983. *The change masters: Innovation and entrepreneurship in the American corporation.* New York: Simon and Schuster.

Katzenbach, J. R., and D. K. Smith. 1993. *The wisdom of teams.* New York: Harper Business.

Kohn, A. 2001. Fighting the tests: A practical guide to rescuing our schools. *Phi Delta Kappan* 82(5): 348–357.

Krijcie, R. V., and D. W. Morgan. 1970. Determining sample size for research activities. *Education and Psychological Measurement 30: 606–610.*

Leithwood, K. 1994. Leadership for school restructuring. *Educational Administration Quarterly* 30(4): 498–518.

Leithwood, K., B. Dart, D. Jantzi, and R. Steinbach. 1993. *Building commitment to change and organizational learning* (Phase 4 final report). Victoria, BC: British Columbia Ministry of Education.

Leithwood, K., L. Leonard, and L. Sharratt. 1998. Conditions fostering organizational learning in schools. *Educational Administration Quarterly* 34(2): 243–276.

Little, J. W. 1990. The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record* 91(4): 509–536.

Lortie, D. C. 1975. *Schoolteacher: A sociological study.* Chicago: University of Chicago Press.

McColskey, W., and N. McMunn. 2000. Strategies for dealing with high-stakes state tests. *Phi Delta Kappan* 82(2): 115–120.

Monti, J., J. York-Barr, R. Kronberg, J. Stevenson, B. Vallejo, and C. Lunders. 1998. *Reflective practice: Creating capacities for school improvement.* Minnesota University, Minneapolis Institute on Community Integration. (ERIC Document Reproduction Service No. ED 427 470).

Owens, T. R. 1979. *Program evaluation skills for busy administrators.* Portland, OR: Northwest Regional Educational Laboratory.

Posavac, E. J., and R. G. Carey. 1997. *Program evaluation: Methods and case studies* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Purkey, S. C., and M. S. Smith. 1983. Effective schools: A review. *Elementary School Journal* 83: 427–452.

———. 1985. School reform: The district policy implications of the effective school literature. *Elementary School Journal* 85: 353–389.

Ravid, R. 1994. *Practical statistics for educators.* Lanham, MD: University Press of America.

Schmoker, M. 1999. *Results: The key to continuous school improvement* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Schön, D. A. 1983. *The reflective practitioner: How professionals think in action.* New York: Basic Books.

Scriven, M. 1967. The methodology of evaluation. In *Curriculum evaluation*, edited by R.E. Stake. American Educational Research Association Monograph Series on Evaluation, No. 1, pp. 39–83. Chicago: Rand McNally.

———. 1991. Beyond formative and summative evaluation. In *Evaluation and education: At quarter century*, edited by M. W. McLaughlin, and D. C. Phillips, pp. 19-64. Chicago: National Society for the Study of Education.

Senge, P. M. 1990. *The fifth discipline: The art and practice of the learning organization.* New York: Doubleday/Currency.

Sergiovanni, T. J. 1984. Leadership and excellence in schooling. *Educational Leadership* 41(5): 4–13.

Shepard, L. A. 2000. The role of assessment in a learning culture. *Educational Researcher* 29(7): 4–14.

Smith, M. L., and G. V. Glass. 1987. *Research and evaluation in the social sciences.* Englewood Cliffs, NJ: Prentice-Hall.

Smith, S. 1999. Leadership training for cultural diversity. *MultiCultural Review* 5(1): 33–38.

Suter, W. N. 1998. *Primer of educational research.* Boston: Allyn and Bacon.

Vaughan, E. D. 1998. *Statistics: Tools for understanding data in the behavioral sciences.* Upper Saddle River, NJ: Prentice-Hall.

Wheatley, M. J. 1994. *Leadership and the new science.* San Francisco: Berrett-Koehler Publications.

Weiss, C. H. 1998. *Evaluation.* Upper Saddle River, NJ: Prentice-Hall.

Wiggins, G. 1989. A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan* 70(9): 703–713.

———. 1994. None of the above. *Executive Educator* 16(7): 14–18.

———. 1996. Embracing accountability. *New Schools, New Communities* 12(2): 4–10.

Wolf, D. 1987/1988. Opening up assessment. *Educational Leadership* 45(4):24–29.

———. 1989. Portfolio assessment: Sampling student work. *Educational Leadership* 46(7): 35–39.

Wolf, D., P. LeMahieu, and J. Eresh. 1992. Good measure: Assessment as a tool of reform. *Educational Leadership* 49(8):8–13.

Worthen, B. R., J. R. Sanders, and J. L. Fitzpatrick. 1997. *Program evaluation: Alternative approaches and practical guidelines.* White Plains, NY: Longman.

INDEX

Accountability, program evaluation and, 5
"Accuracy" standard, 93n2
Achievement tests, program assessment and, 78
Acknowledgment page, in evaluation report, 149
Actions, power of positive attitude on, 22–23
Analysis
  of attitudinal data, 145
  of data to assess program's impact, 133–146
  effect size approach for data analysis, 135–137
  of observational data, 142–144
  one- or two-group comparisons, 141
  of portfolios, 144–145
  of qualitative data, 146
  quantitative data, 134
  single-group comparison, 140
  test of significance, 134–135
  two-group comparison, 137–140
Assessment. *See also* Program evaluation
  authentic, 86–87
  authentic, sources for, 185–186
  high stakes, xvii
  of program processes, 7–9
  traditional linked with authentic, 89–90
Attitude(s)
  evaluation and measurement of student, 107
  power of on action, 23
  principal's, toward program assessment, 174–176
  promoting positive, 22
Attitude dimension, of reflective practice, 21
Attitudinal data analysis, 145
Attrition, as threat to internal validity, 37
Authentic assessment
  meaningful learning and, 87–88
  vs. "mindless" testing, 88
  perspectives on, 86–87
  representative sources for, 185–186

Classroom observations
  for evaluation, 117–121
  procedures for conducting, 118
Classroom rating scale, 119–121
Classroom structure, systems thinking and compartmentalized, 24
Class selection, for sampling in evaluation information collection, 104–106
Cluster, in sampling, 104
Cluster sampling, 104
Collaboration. *See also* Evaluation team; Team entries
  collective problem solving and, 26
  evaluation team, teaching, and, 25
  need for, 68
  program evaluation through, 61–73
Collaborative learning culture, evaluation team formative assessment and, 168
Collective intelligence, school improvement and, 184n3
Collective judgment, of evaluation team, 62
Collective problem solving, 26
Collective search, facilitated through evaluation project, 21
Communication, establishing evaluation team and, 66
Compartmentalized structures, systems thinking and, 24, 30n2
Competency dimension, of reflective practice, 21
Comprehensive evaluation, 78–79
Computer programs, as short-term fix for standardized testing skills, 27

BEST COPY AVAILABLE

Confidence, achievement and, 183, 184n5
Confidentiality and anonymity, in program evaluation, 122–125, 131n9
Confirmative (confirm) function, of program evaluation, 1–2, 66, 156
Confounding, 34–35, 41, 43n4
Consensus, for final evaluation questions selection, 102–103
Content validity, of tests, 82–84, 85
Contextual factors, affecting program's outcomes, 147–148
Control group, 39. *See also* Experimental program evaluation
  evaluation information collection and, 125, 127–129
  nonrandomized pretest-posttest, 52–53
  one-group time series, 53–55
  randomized pretest-posttest, 51–52
  use of, 46–47
Cost-benefit and cost-effectiveness analyses, to determine curriculum efficiency, 101
Costs. *See also* Financial resources
  of program, 98
Creativity, meaningful learning and, 87–88
Criterion-referenced tests, using to profile student achievement, 139–140, 150n4
Culture of inquiry, 28
Current perspective. *See also* Non-experimental program evaluation
  of program evaluation, 32
Curriculum. *See also* Program evaluation
  cyclical process of improving, 9–11
  determining efficiency through evaluation questions, 100–101
  developmental approach to improving, 31–41
  evaluation and, 1
  as "work in progress," 11–12
Cyclical process, of improving curriculum, 9–11

Data. *See also* Evaluation data
  qualitative, 77–78
  quantitative, 77, 78
Data collection
  methods in evaluation process, 107–109
  parts of, 93n3
  process for collecting valid data, 80–82
Demoralization, as threat to internal validity, 38
Dependent variable (outcome variable), in program evaluation, 33
DeRoche's principles of evaluation, 6
Describing the program (evaluation process). See Program description (evaluation process)
Diagnostic (diagnoses) function, of program evaluation, 1–2, 66, 156–157
Dialogues, on ways to improve student achievement, 26
Diversity of knowledge and skills, evaluation team and, 62
Draft version, of evaluation report, 149

Effectiveness
  compared to efficiency, xvi
  evaluating program's, 146–148
  indicators of to gauge professional performance levels, 176
Effectiveness dimension of evaluation, evaluation questions and, 99–100
Effect size, 135–137
  calculating for subgroups of students, 137, 138
  criteria for interpreting, 136
  formula for, 135
  percentile shift and, 135–136
  ranges of, and corresponding percentile shifts, 187–189
Efficiency
  compared to effectiveness, xvi
  determining curriculum, 100–101
  evaluating program's, 148
Effort, excellence and, 11
ERIC database, 107–108
Ethical issues
  in program evaluation, 122–125, 131n9
  valid data collection process, evaluation team, and, 82
Evaluation
  comprehensiveness of, 78–79
  DeRoche's principles of, 6
  formative and summative, 6–7, 7–9
  of program processes, xv–xvi
Evaluation data
  collecting process, 81–82
  collecting valid data, 80–81
  quantitative and qualitative, 77–78
  reliability of, 80, 84–86
  tests, content validity, and, 82–84

validity of, 79–80, 80–84
Evaluation information collection, 103–129
classroom observations and, 117–121
follow-up questions for, 110
guidelines for, 125–129
methods for, 107–109
preliminary considerations, 103
program needs assessment and, 110–117
random class selection for sampling, 104–106
using a sample, 104
Evaluation Plan, 122, 123
Evaluation process, 95–149
phase 1—describing the program, 96–98
phase 2—providing direction for the evaluation, 99–103
phase 3—obtaining information to answer evaluation questions, 103–129
phase 4—analyzing data to assess program's impact, 133–146
phase 5—evaluating program's effectiveness and offering recommendations for future development, 146–148
phase 6—writing evaluation report, 149, 151–160
Evaluation project. *See also* Evaluation process; Program evaluation
assessing, 163–171
collective search facilitated through, 21
evaluation questions for site-based, 102
formative metaevaluation, 164
judging, 169–170
monitoring, 165–167
organizing procedures for, 122, 123–124
phases of, 20
summative metaevaluation, 168
Evaluation project assessment questionnaire, 169–170
Evaluation questions
formulation of, 99–102
obtaining consensus on final selection of, 102–103
obtaining information to answer, 103–129
presentation to stakeholders, 102
Evaluation report, 125
abstract, 153
acknowledgment page in, 149
acknowledgments, 153
appendices, 156
checklist for, 158–160
conclusions, 155–156
contents of, 151
draft version, 149
evaluation questions, 154
evaluation results and interpretation, 155
final report, 149
procedures for collecting and analyzing evaluation information, 154–155
program description, 153
purpose of evaluation, 153
reflective practice and metaevaluation, 171
as representation of confirmative and diagnostic functions of program evaluation, 156–157
title page, 152
Evaluation study, process dimension of, 90
Evaluation team, 3, 4, 14n2. *See also* Evaluation process
administrator's role in, 69
advantages of using, 61–63
collaboration and, 25
collective judgment and, 62
composition of, 63–64
conducting and setting tone for meetings, 69–71
data collection issues to discuss between program faculty and, 82, 83
election of leader and recorder of activities, 69
factors for principals to consider in establishing, 64–65
guidelines for assembling, 65–68
input of members, 71
knowledge and skills diversity and, 62
learning organization environment and, 25
meeting agenda for, 71
meetings of, 69–72
preliminary activities of, 69
in program improvement cycle, 10
qualifications and responsibilities of team leader, 68
questions for formative assessment of, 168
responsibilities of, 68
shared responsibility and, 63

BEST COPY AVAILABLE

systems mentality and, 23
valid data collection process and, 81–82
varied perspectives of, 62–63
Evaluation Time Log, 122, 124
Evaluators, determination of program success by, 134
Excellence
effort and, 11
in evaluation report, 157
Experiment. See also Experimental designs
achieving internal validity in an experiment: confirming if the experimental treatment itself was effective, 48
building control into, 46–47
external validity and, 56–58
internal validity and, 35–40, 45–46
matching and random assignment in, 49–50
planning, 47
random assignment and, 48–49, 50–51
Experimental designs
decision to use one or two groups and, 56
features of, 57
nonrandomized pretest-posttest control group, 52–53
one-group time series, 53–55
randomized pretest-posttest control group, 51–52
Experimental groups, evaluation information collection and, 125, 127–129
Experimental program evaluation, 31, 32–34, 45–58
compared with nonexperimental methods, 34
confounding and, 34–35, 41
dependent variable (outcome variable) and, 33
independent variable and, 33
terms and principles of, 33
threats to internal validity and, 35–40
Experimenter effect, as threat to internal validity, 38–39
External validity, 56–58
Extraneous variables
confounding and, 34
random assignment and, 51
use of control group and, 46–47

Faculty
leadership initiatives of, 177
leadership in program evaluation, 3–4
Fairness, in evaluation report, 157, 161n4
Feedback
from draft version of evaluation report, 149
evaluation and, 30n6
in program improvement cycle, 10
Final evaluation report, 149
Financial resources, efficiency dimension of evaluation and, 148
Focus groups
to address evaluation questions, 117
interviews and qualitative data analysis, 146
sampling and, 106–107
Follow-up questions, in evaluation information collection, 110
Formative assessment, of evaluation team, questions for, 168
Formative evaluation, 6–9
metaevaluation and, 163
in program improvement cycle, 10
reflective practice and, 28
Formative metaevaluation, 164
Formative methodology, 8
Formative purposes, observational data analysis and, 143
Functions, of program evaluation, 1–2
Future perspective. *See also* Experimental program evaluation
of program evaluation, 32

Goals
and activities, for evaluation team, 67
of program to be evaluated, 97
Growth tendencies, unleashing, 11–12

High stakes assessment, xvii
High stakes testing, minimizing risks associated with, 25–27
History threat, internal validity and, 35–37
Holistic thinking, program assessments and, 25

Independent variable, in program evaluation, 33, 46
Individualism, systems thinking and, 24
Information. See Data entries; Evaluation information collection; Evaluation questions

Informed consent, in program evaluation, 125
Innovative program, experimental evaluation and, 32–33
Instructional activities
  creativity and, 87–88, 89
  of program, 98
Instruction techniques, exchanging effective, 26
Instrument, defined, 93n1
Instrumentation, internal validity and, 39–40
Internal validity, 35–40
  attrition threat to, 37
  demoralization threat to, 38
  experimenter effect and, 38–39
  history threat and, 35–37
  instrumentation and, 39–40
  John Henry effect and, 38
  maturation and, 39
  pretesting and, 40
  review, 45–46
  selection threat and, 37
  various threats to that can boost an innovative program's effectiveness, 41
Intervention
  experimental evaluation and, 43n1
  in program improvement cycle, 11
Interviews
  to address evaluation questions, 117
  for data collection, 108–109
  qualitative data analysis and focus group, 146
  questions for staff, 128
  with students, 109
Isolation, in teaching, 24
Item analysis, for observational data, 144

John Henry effect, as threat to internal validity, 38

Kaizen, spirit of, adoption by learning organization, 17–18
Knowledge and skills diversity, of evaluation team, 62

Leadership
  in program evaluation, 3–4
  qualities in principal's, 176–177
  staff initiatives, 177
  transformational, 19–20
Leadership role
  of principal in program evaluation, 173–182
  in program evaluation, xviii
Leadership strategies
  communicating positive outlook toward school improvement efforts as ultimate, 22
  to ensure effective program evaluation, xvi
Learning
  authentic assessment and meaningful, 87–88
  stimulus for organizational, 21
Learning culture, 19
Learning organization
  adoption of spirit kaizen by, 17–18
  defined, 18
  formative and summative methods integration in, 28
  linking parts of for improved performance, 23–25
  program evaluation contributions to, 17–28
  reflective practice and, 20–23
  school as, 18–20
  team learning and, 24
"Level playing field" metaphor, experimental and control groups and, 47
Likert scale, 108, 127

Management activities, in program, 98
Matching, combining with random assignment, 49–50
Material resources, for program, 98
Maturation, internal validity and, 39
Meaningful progress, program evaluation and, 134
Meaningful value, as indicator for evaluating program outcomes, 150n1
Measurement of program outcomes, 77–91
  comprehensive evaluation and, 78–79
  meaningful learning, creativity, and, 87–90
  multidimensions of, 90–91
  quantitative and qualitative data and, 77–78
  teacher opinions and, 78–79
Meetings, of evaluation team
  agenda for, 71
  enthusiasm for, 72
  member input at, 71

BEST COPY AVAILABLE

setting tone for, 70–71
starting of, 71
suggestions for team leader, 69–70
Metaevaluation
formative, 164
purposes of, 163–164
reflective practice and, 171
summative, 168
Moderator. *See* Team leader
Modified program, experimental evaluation and, 32–33

Needs assessment, program, 110–117
Nominations, for evaluation team, 66
Nonexperimental program evaluation, 31, 32
compared with experimental methods, 34
independent and dependent variables in, 33
Nonrandomized pretest-posttest control group, 52–53. *See also* Control group
Nonstandardized tests, determining validity of, 84, 85

Objectives, of program to be evaluated, 97
Objectivity, program assessment and, 161n3
Observational data analysis
formative and summative purposes of, 143
item analysis for, 144
procedure for, 142
Observations, classroom, for evaluation, 117–121
One-group time-series design, 53–55
questions for assessing, 54–55
sample schedule for, 54
One- or two-group analysis, using report card data, 141
Open-ended questions
for assessing one-group time-series design, 54–55
qualitative data analysis and, 146
Optimism, achievement and, 183
Organizational functioning, effectiveness and efficiency and, xvi
Organizational improvement, systems approach to, 11
Organizational learning
evaluation, feedback, and, 30n6
stimulus for, 21
Organization development, program evaluation and, xvi, 4–5
Outcomes (program). *See* Program outcomes
Outcome variable (outcome). *See* Dependent variable (outcome variable)

Parents, survey questionnaires for data collection, 108–109
Percentile shifts
determining from effect size, 135–136
ranges of effect size and corresponding, 187–189
Performance assessment, authentic assessment and, 86–87
Personal mastery, in learning organization, 181
Personnel, of program, 98. *See also* Faculty; Program faculty
Perspectives
on authentic assessment, 86–87
of evaluation team, 62–63
Portfolio analysis, 144–145
Portfolios
authentic assessment and, 86–87
displaying creative work in, 87–88
Positive attitudes, promotion of in learning organization, 22
Practical significance. *See also* Effect size
program's impact and determination of, 134
Practice standardized tests, as short-term fix, 27
Presentism, teamwork and, 24
Pretesting, internal validity and, 40
Principals. *See also* Evaluation team
active intervention in evaluation, 174
attitude toward program assessment, 174–176
collection of valid data and, 81
establishing evaluation team guidelines for, 65–68
leadership qualities of, 176–177
leadership role in program evaluation, 3–4, 173–182
learning organization environment and, 25
and obstacles in establishing evaluation team, 64–65
role at first meeting with evaluation group, 69
role in program evaluation, xviii

self-assessment of role in program evaluation, 178, 179–180
staff leadership initiatives and, 177
transformational leadership of, 19–20
Principles of evaluation, of DeRoche, 6
Process dimension, of evaluation study, 90
"Process" information, methodology for, xvi
Professional development and growth
program evaluation and, xvi
site-based program evaluation and, 4
Professionalism
assessment of school personnel's work and, xvii
in evaluation report, 157
Professional performance, indicators of program effectiveness used to gauge, 176
Program
context factors and, 147–148
defined, 14n1
efficiency of, 148
evaluating effectiveness and offering recommendations for future development, 146–148
linking evaluation with, 9–11
as "work in progress," 11–12
Program description (evaluation process)
goals and objectives, 97
operation, 97–98
rationale for program, 96–98
writing, 96
Program evaluation, xv. *See also* Evaluation process; Evaluation project; Learning organization; Program outcomes
benefits of site-based, 4–5
collaboration and, 61–73
compatible methods of, 31–34
comprehensiveness of, 78–79
confounding and, 34–35, 41
ethical aspects of, 122–125
experimental, 31, 32–34
experimental approach to, 45–58
fair and complete assessment in identifying strengths/weaknesses, 75n3
as feature of school as learning organization, 19
formative and summative approaches to, 6–9
functions of, 1–2
goals of, 5–6
holistic thinking and, 25
nonexperimental, 31, 32, 34
ongoing basis of, 3
personal/human side of, 81
perspectives of, 32
perspectives on, 1–12
phases of evaluation project, 20
principals and faculty leadership in, 3–4
school improvement and, 1–4
school leader's role in, xviii
Program faculty. *See also* Faculty; Teachers
data collection issues, evaluation team, and, 82, 83
interviews of for data collection, 108
Program improvement cycle, 9–11
Program needs assessment, 110–117
Program operation in evaluation process
characteristics of participants, 97
costs, 98
instructional activities, 98
management activities, 98
material resources, 98
program personnel, 98
schedule, 98
selection of participants, 97
Program outcomes
measurement of, 77–91
value of evaluating in comprehensive way, 175
Program processes
analyzing, xv–xvi
assessing, 7–9
formative evaluation and, 7–9
Program staff. *See* Program faculty

Qualitative data, 77–78, 129
Qualitative data analysis, 146
Quantitative data, 77, 78, 129
Quantitative data analyses, 133–134
Questionnaires. *See also* Program needs assessment; Survey questionnaires
attitude, 145
evaluation project assessment, 169–170
Questions. *See also* Evaluation questions
follow-up in evaluation information collection, 110
for formative assessment of evaluation team, 168
for staff interviews, 128
for student questionnaires, 128
in two-group design, 126–127
when forming evaluation team, 67

BEST COPY AVAILABLE

Random assignment
as experimental control method, 48–49
limitations on use of, 50–51
used with matching, 49–50
Randomized pretest-posttest control group, 51–52. *See also* Control group
Rating scale, classroom, 119–121
Rationale, for evaluation team, 66
Recommendations, for program improvement, 80
Reflective openness, 21
Reflective practice, 20–23, 28, 30n6
areas in need of improvement and, 176
attitude dimension of, 21
competency dimension of, 21
defined, 21
evaluation report and, 157
leadership roles, program evaluation, and, xvi
metaevaluation and, 171
Re-funding proposal, program improvement and, 130n4
Reliability
of evaluation data, 80
of tests, 84–86
Remediation, diagnoses and, 2
Report, for evaluation team, 67. *See also* Evaluation report
Report card data, using in one- or two-group analysis, 141
Response set, questionnaires and, 128
Responsibilities, of evaluation team, 68

Sampling, 131n6
cluster, 104
excluding stakeholders from, 131n7
focus groups and, 106–107
population size and required sample size for, 105
systematic, 106
using in evaluation information collection, 104–107
Schedule, for program, 98
School. *See also* Learning organization
efficiency dimension of evaluation and resources of, 148
as learning organization, 18–20

School improvement
collective intelligence and, 184n3
endless nature of, 31–32
program evaluation and, 1–4
staff leadership initiatives and, 177
School leadership. *See* Leadership
School personnel, assessment of work as professionalism indicator, xvii. *See also* Faculty; Teachers
School-wide leadership, 177
Search, facilitated through evaluation project, 21
Selection
of program participants, 97
as threat to internal validity, 47
Self-assessment, of principal's role in program evaluation, 178, 179–180
Self-improvement
in learning organization, 181–182
positive perception of testing and, 181
"Shared leadership," xi
Shared responsibility, of evaluation team, 63
Single-group comparison, to determine improvement, 140
Site-based evaluation project, 4–5
evaluation questions for, 102
Size, of evaluation team, 67
Sources, for authentic assessment, 185–186
Specialness, motivating effect of, 25, 30n3
Staff. *See* Faculty; Principals; Teachers
Stakeholders
evaluation questions presented to, 102
excluded from sampling, 131n7
included in evaluation, 131n8
Standard deviation, 150n2
Standardized tests
creating positive perception of, 178–181
creativity, meaningful learning, and, 89–90
as measurement of organizational performance, 174–175
as measurement of program's success, 175
minimizing risks associated with, 25–27
practice tests, 27
using to determine school success, xvii
using to measure extent of differences from pre- to posttests, 139
Statistical analyses
computer program and, 134
effect size and, 135–137
test of significance, 134–135
Student achievement, profile of using criterion-referenced tests, 139–140

Student outcomes, analyzing, xv–xvi
Student performance, xv
raising through systems thinking approach, 25–28
site-based program evaluation and, 4
Students
developing test-taking skills of, 26
interviews for data collection, 109
measurement of attitudes for program evaluation, 107–108
practice standardized tests for, 27
survey questionnaires for data collection, 109
Subgroups, determining effectiveness of experimental treatment by comparing, 137, 138
Success, need for, xv
Summative evaluation, 6–9
metaevaluation and, 163
in program improvement cycle, 9, 10
reflective practice and, 28
Summative metaevaluation, 168
Summative purposes, observational data analysis and, 144
Support, achievement and, 184

Survey questionnaires
for experimental and control groups, 127–129
for parents, 108–109
qualitative data analysis and, 146
response set and, 128
reverse scoring of, 129
for students, 109
Systematic sampling, 106
Systems approach, to organizational improvement, 11
Systems thinking
compartmentalized classroom structure and, 24
evaluation team and, 23, 63
increasing student achievement through, 25–28

Teacher-made tests, creativity, meaningful learning, and, 89–90
Teacher opinions, evaluation data and, 78–79
Teachers. *See also* Program faculty
collective problem solving and collaboration, 26
communication across grade levels to improve student achievement, 26
on evaluation team, 63
as obstacles to evaluation, 64–65
Teaching
collaboration and, 25
egocentric nature of, 64, 75n6
isolation in, 24
Team approach, value of, 75n1
Team leader. *See also* Evaluation team
dynamics of group processes and, 64
election of, 69
qualifications and responsibilities of, 68
setting tone for meetings and, 70–71
suggestions for conducting meetings, 69–70
Team learning, 24
Teamwork. *See also* Collaboration; Evaluation team
presentism and, 24
Technical approach, to determine curriculum efficiency, 101
Technical dimension, of evaluation, 81, 94n4
Tenure, of evaluation team, 67
Test(s). *See also* Achievement tests; High stakes testing; Standardized tests
content validity of, 82–84, 85
creativity, meaningful learning, and, 89–90
criterion-referenced, 139–140, 150n4
emphasis on results of, xv
reliability of, 84–86
using standardized to determine school success, xvii
Testing, creating positive perception of, 178–182
Test of significance, 134–135
Test performance, improving, 87–90
Test-taking skills, developing in students, 26
Time frame, for evaluation team, 67, 75n8
Time resources, efficiency dimension of evaluation and, 148
Transformational leadership, 19–20
Transforming leadership, 19, 30n1
Two-group comparisons, 137–140
using percentages to measure differences in, 139
Two-group design, of program evaluation, 126–127

Ultimate leadership strategy, communicating positive outlook toward school improvement efforts as, 22
Unobtrusive measures, for data collection, 107

Unplanned events, in experimental program evaluation. See History threat

Validity
content validity of tests, 82–84
of evaluation data, 79–80, 80–84
internal, 35–40
Value judgments, program effectiveness and, 6
Variance, experimental program evaluation and, 34–35
Varied perspectives, of evaluation team, 62–63
Volunteers, for evaluation team, 66, 71

Work, providing meaningful and challenging, 27

EVALUATING PROGRAMS TO INCREASE STUDENT ACHIEVEMENT

LEADERSHIP STRATEGIES

"Whether school leaders need the foundation of knowledge and attitudes applicable to conducting an evaluation or a step-by-step procedure to begin an evaluation project, this book should be considered their guide."

—Cozette Buckney, Ed.D.
Liaison to the Board President/CEO
Chicago Public Schools

With the ever-present call to increase student achievement comes the challenge of evaluating the programs that will accomplish that goal. Dr. Martin Jason has created a framework for such evaluation and embedded strategies for leaders to use to assess programs and affect change in their schools. This book is a practical guide for novice leaders and a true resource for seasoned educators.

MARTIN H. JASON, PH.D.

Dr. Jason is Professor of Education in the Educational Leadership and Organizational Change Program at Roosevelt University. Dr. Jason is also Director for the Educational Administration and Supervision Program and teaches courses at the master's as well as doctoral levels, including educational research methods, data analysis, program and personnel evaluation, and educational leadership.

Dr. Jason has presented papers at a number of educational leadership conferences, including the Phi Delta Kappa International Conference on Effective Schools and the annual conference of the Mid-Western Educational Research Association. He has also published a number of articles in scholarly texts and journals, including *High School Journal* and *Research for Educational Reform.*

BEST COPY AVAILABLE

SkyLight Professional Development

Visit our Web site at
www.skylightedu.com

ISBN 1-57517-815-X
90000
9 781575 178158

218

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

ERIC Clearinghouse on Elementary
and Early Childhood Education
and the National Parent Information Network

29 Children's Research Center
51 Gerty Drive
Champaign, IL 61820-7469 USA

July 28, 2003

ERIC DOCUMENT TRANSFER FORM

To:

ERIC/AE
UNIV. OF MD-COLLEGE PARK
1129 SCHRIVER HALL
COLLEGE PARK, MD 20742

If you decide to include this document in the database, please use PS's Level 2B blanket.

From:

ERIC/PS
51 Gerty Dr.
CHAMPAIGN, IL 61820-7469
Phone: (217) 333-1386
Fax: (217) 333-3767

Date Received: 7/25/03

The enclosed document appears to be more within your scope than ours.

The document was obtained by:

- ✓ solicited contribution
- ____ unsolicited contribution
- ____ staff contribution
- ____ ERIC facility

All relevant information is attached and the author has been notified of the transfer.

A. Smith
ERIC/PS Acquisitions